

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 24, 2005, 01:32:23 ; Search time 165 Seconds  
(without alignments)  
1036.049 Million cell updates/sec

Title: US-10-016-768A-8  
Perfect score: 2250  
Sequence: 1 MKQMIRQFAIEYISKSGTKQ.....GLYNMTDSTGSCNKSXPV 442

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_16Dec04:.\*  
1: Geneseqp1980s:.\*  
2: Geneseqp1990s:.\*  
3: Geneseqp2000s:.\*  
4: Geneseqp2001s:.\*  
5: Geneseqp2002s:.\*  
6: Geneseqp2003as:.\*  
7: Geneseqp2003bs:.\*  
8: Geneseqp2004s:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2250	100.0	442	5	Aae24371 Human E93
2	1913.5	85.0	630	4	Abg17942 Novel hum
3	1754	78.0	356	5	Adh48740 NOV10 pro
4	1411	62.7	277	8	Abg60261 Human gen
5	1090	48.4	213	6	Ada54381 Human pro
6	502	22.3	311	7	Adm04876 Human pro
7	499.5	22.2	433	8	Adn35086 Nuclear r
8	499.5	22.2	572	7	Adel1284 Novel hum
9	499.5	22.2	572	7	Adj69960 Human hea
10	294.5	13.1	104	5	Abp32451 Human ORF
11	273	12.1	53	5	Aae24592 Human E93
12	233.5	10.4	54	5	Aae24593 Fish E93
13	229	10.2	53	5	Aae24594 Mouse E93
14	200.5	8.9	1140	4	Abb71145 Drosophil
15	200.5	8.9	1165	5	Aae24372 Fruit fly
16	165	7.3	53	5	Aae24370 Fruit fly
17	163	7.2	53	5	Aae24595 Nematode
18	133.5	5.9	1109	8	Adr70231 Rabbit GM
19	133.5	5.9	3616	8	Adf45530 Chicken A
20	132.5	5.9	848	5	Abg69087 Botulinum
21	132.5	5.9	849	8	Adn11042 Clostridi
22	131.5	5.8	870	8	Abg59193 Human gen
23	131.5	5.8	2288	8	Adk60224 Angiogene
24	131.5	5.8	2288	8	Adk60525 Angiogene
25	131.5	5.8	2288	8	Adp73148 Angiogene

26	131.5	5.8	2375	7	ADJ68471 Human hea
27	131.5	5.8	2492	8	ADK60200 Angiogene
28	131.5	5.8	2492	8	ADK60501 Angiogene
29	131.5	5.8	2492	8	ADP73124 Angiogene
30	131.5	5.8	2492	8	ADQ89534 Human ATR
31	130	5.8	3248	2	AAR99795 Kinetoch
32	127	5.6	533	5	AAR82977 S. cerevi
33	127	5.6	534	6	ABR53199 Protein s
34	127	5.6	534	7	ADK63276 Disease t
35	126.5	5.6	750	8	ADL05858 M. catarr
36	126	5.6	433	8	ADN19824 Bacterial
37	125.5	5.6	2285	4	ABG63057 Drosophil
38	125	5.6	2759	6	AAO16418 Human nuc
39	124.5	5.5	972	5	AAY076761 Plasmodi
40	124.5	5.5	2954	2	AAO16332 Amino aci
41	124.5	5.5	3187	7	ADK56031 Rat Prote
42	124.5	5.5	3187	7	ADK56035 Rat Prote
43	123.5	5.5	1827	8	ADN22715 Bacterial
44	123	5.5	454	5	AAU91510 Outer sur
45	123	5.5	1017	7	ADJ68352 Human hea

ALIGNMENTS

RESULT 1  
AAE24371  
ID AAE24371 standard; protein; 442 AA.  
XX  
AC AAE24371;  
XX  
DT 04-OCT-2002 (first entry)  
XX  
DE Human E93 programmed cell death modulating protein.

XX Human; cancer; programmed cell death modulating protein; adenocarcinoma;  
cellular apoptosis; leukaemia; acquired immunodeficiency syndrome; AIDS;  
neurodegenerative disease; Alzheimer's disease; retinitis pigmentosa;  
Parkinson's disease; myelodysplastic syndrome; cerebellar degeneration;  
aplastic anaemia; ischaemic injury; myocardial infarction; stroke;  
reflex anisokoria; toxin-induced disease; genetic immunodeficiency;  
vaccine; gene therapy; lymphoma; cytostatic; melanoma; neuroprotective;  
myeloma; neutropenic; vasotropic; immunostimulant; cerebroprotective;  
cardiac; E93 protein.  
XX Homo sapiens.

XX Key Location/Qualifiers  
PH Domain 353..405  
/note= "Conserved domain"  
FT WO200234882-A2.  
PN 02-MAY-2002.  
XX  
PD 29-OCT-2001; 2001WO-US048053.  
XX 27-OCT-2000; 2000US-0243865P.  
XX (UYMA-) UNIV MARYLAND BIOTECHNOLOGY INST.

XX Baehrecke EH;  
XX WPI; 2002-479717/51.  
XX Novel programmed cell death modulating proteins, useful for treating or  
preventing disorders associated with abnormal cell proliferation and  
apoptosis such as cancer, stroke, Parkinson's disease, myocardial  
infarction.  
XX Claim 1; Fig 4; 89pp; English.  
XX The present invention relates to novel programmed cell death modulating

CC proteins and polynucleotides encoding such proteins. Sequences of the  
CC invention are useful to screen potential cellular apoptosis inhibiting  
CC compounds to determine their use as therapeutic agents for treatment of  
CC diseases associated with increased programmed cell death. They are also  
CC useful for treating or preventing disorders associated with decrease in  
CC apoptosis. Programmed cell death modulating sequences are useful for  
CC treating or preventing cancer e.g. adenocarcinoma, leukaemia, lymphoma,  
CC melanoma, myeloma. Inhibition of the activity of the sequences of the  
CC invention are useful for treating disorders associated with increase in  
CC cell death or apoptosis such as acquired immunodeficiency syndrome  
CC (AIDS), neurodegenerative diseases (e.g., Alzheimer's disease, retinitis  
CC pigmentosa, Parkinson's disease and cerebellar degeneration), ischaemic  
CC injuries (e.g., myocardial infarction, stroke, reperfusion injury),  
CC myelodysplastic syndromes (e.g., aplastic anaemia), toxin-induced  
CC diseases and other infectious or genetic immunodeficiencies. Sequences of  
CC the invention are used as vaccines and in gene therapy. The present  
CC sequence is human E93 programmed cell death modulating protein  
XX  
SQ Sequence 442 AA;  
..Query Match 100.0%; Score 2250; DB 5; Length 442;  
Best Local Similarity 100.0%; Pred. No. 9.9e-175;  
Matches 442; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MKKIROFALEYISKSGKTOENRNGSTGPIVCKSIQMQAENSLOEQEGPLDLTVNRM 60  
Db 1 MKKIROFALEYISKSGKTOENRNGSTGPIVCKSIQMQAENSLOEQEGPLDLTVNRM 60  
Qy 61 QEQNTQGGVLDLSTTKTSIKSESSICDPSSNSVAGRLHNRNEDYVERSAEFADGILL 120  
Db 61 QEQNTQGGVLDLSTTKTSIKSESSICDPSSNSVAGRLHNRNEDYVERSAEFADGILL 120  
Qy 121 SKALDIOQSGALDINKAGILYGIPOKTLHLHLALPAGKPAKFNKTRPHDSYSYKDSK 180  
Db 121 SKALDIOQSGALDINKAGILYGIPOKTLHLHLALPAGKPAKFNKTRPHDSYSYKDSK 180  
Qy 181 ETCVAVLOKVALWARAQARTEKSKNLNLETSEIKFPTASTYLHQLTLQKMTVTFEKNES 240  
Db 181 ETCVAVLOKVALWARAQARTEKSKNLNLETSEIKFPTASTYLHQLTLQKMTVTFEKNES 240  
Qy 241 LQYETSNPTVOLKIPOLRVSSVSKSQPDGSLDLVMTQVSKTSVLEGSALQKLNILPK 300  
Db 241 LQYETSNPTVOLKIPOLRVSSVSKSQPDGSLDLVMTQVSKTSVLEGSALQKLNILPK 300  
Qy 301 QNKIECSGPTVTHSSVDFYFLHGDLSPLCLNSKNGTVDTGTSNTEDGLDRKDSKOPKRG 360  
Db 301 QNKIECSGPTVTHSSVDFYFLHGDLSPLCLNSKNGTVDTGTSNTEDGLDRKDSKOPKRG 360  
Qy 361 RYROYDHEIMEEALAMVMSGKMSVSKAQGIYGVPHSTLEYKVKERSGTLKTPPKKRLRP 420  
Db 361 RYROYDHEIMEEALAMVMSGKMSVSKAQGIYGVPHSTLEYKVKERSGTLKTPPKKRLRP 420  
Qy 421 DTGLYNTDSTGSGCKNSKPV 442  
Db 421 DTGLYNTDSTGSGCKNSKPV 442  
RESULT 2  
ABGI17942  
ID ABGI17942 standard; protein; 630 AA.  
XX  
XX ABGI17942;  
XX  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #17933.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.

XX PD 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US008631.  
XX  
XX 31-MAR-2000; 2000US-00540217.  
XX 23-AUG-2000; 2000US-00649167.  
XX (HYSE-) HYSEQ INC.  
XX  
XX Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI; 2001-639362/73.  
XX N-PSDB; AAS82129.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess  
XX biodiversity.  
PS Claim 20; SEQ ID NO 48301; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
XX sequences. (I) is useful as hybridisation probes, polymerase chain  
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
XX and in recombinant production of (II). The polynucleotides are also used  
XX in diagnostics as expressed sequence tags for identifying expressed  
XX genes. (I) is useful in gene therapy techniques to restore normal  
XX activity of (II) or to treat disease states involving (II). (II) is  
XX useful for generating antibodies against it, detecting or quantitating a  
XX polypeptide in tissue, as molecular weight markers and as a food  
XX supplement. (II) and its binding partners are useful in medical imaging  
XX of sites expressing (II). (I) and (II) are useful for treating disorders  
XX involving aberrant protein expression or biological activities in  
XX polypeptide and polynucleotide sequences have application in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
XX amino acid sequences of the invention. Note: The sequence data for this  
XX patent did not appear in the printed specification, but was obtained in  
XX electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 630 AA;  
Query Match 85.0%; Score 1913.5; DB 4; Length 630;  
Best Local Similarity 84.1%; Pred. No. 4.7e-147;  
Matches 392; Conservative 11; Mismatches 30; Indels 33; Gaps 4;  
Qy 10 IEVYKSGTKQEN-----RNGSTGPSIVCKSIQMQAENSLOEQE--- 48  
Db 165 MELLSDHDKKVENKIIQTRFRKQETLFAMRNSSDPMFRQSLOIKRELASLDENFTRK 224  
Qy 49 ---QEGPLDLTVNR-----MOEQNTQOG---DGVLDTLSTKTSIKSESSICDPSSNS 96  
Db 225 KYTESKRRLQTNNEISSDDKGEFHEQPNWDGVLDTLSTKTSIKSESSICDPSSNS 284  
Qy 97 VAGRLHNRNEDYVERSAEFAFGLLSKALXDIQSGALDINKAGILYGIPOKTLHLHLALP 156  
Db 285 VAGRLHNRNEDYVERSAEFAFGLLSKALXDIQSGALDINKAGILYGIPOKTLHLHLALP 344  
Qy 157 AGKPASFKNKTDRPHDSYSYKDSKETCAVLQKVALWARAQARTEKSKNLNLETSEIKFP 216  
Db 345 AGKPASFKNKTDRPHDSYSYKDSKETCAVLQKVALWARAQARTEKSKNLNLETSEIKFP 404  
Qy 217 TASTYLHQLTLQKMTVTFEKNESLOYETSNPTVOLKIPOLRVSSVSKSQPDGSGLLDVM 276  
Db 405 TASTYLHQLTLQKMTVTFEKNESLOYETSNPTVOLKIPOLRVSSVSKSQPDGSGLLDVM 464  
Qy 277 YQVSKTSSVLEGSALQKLNILPKQNKIECSGPTVTHSSVDFYFLHGDLSPLCLNSKNGTV 336  
Db 465 YQVSKTSSVLEGSALQKLNILPKQNKIECSGPTVTHSSVDFYFLHGDLSPLCLNSKNGTV 524

QY 337 DGTSENTEGLDRKDSKQPRKGRYQYDHEIMEEAIAMVMSGKMSVSKAQGIYGVPHS 396  
DB 525 DGTSENTEGIDRKDSKQPRKGRYQYDHEIMEEAIAMVMSGKMSVSKAQGIYGVPHS 584  
QY 397 TLEYKVKERSGTLKTPPKKKLRLPDTGLYNMTDSGTGCKNSKSPV 442  
DB 585 TLEYKVKERSGTLKTPPKKKLRLPDTGLYNMTDSGTGCKNSKSPV 630

RESULT 3  
ADH48740  
ID ADH48740 standard; protein; 356 AA.  
AC ADH48740;

XX 25-MAR-2004 (first entry)  
DT NOV10 protein sequence, SEQ ID 24.  
DE  
XX Human; NOVX; atherosclerosis; hypertension; obesity; cancer; cytostatic;  
KW hypotensive; antiarteriosclerotic; anorectic; gene therapy; NOV10;  
KW T01C1.3-like protein; chromosome 4.  
XX  
OS Homo sapiens.  
XX  
PN WO200268652-A2.  
XX  
PD 06-SEP-2002.  
XX  
PF 26-FEB-2002; 2002WO-US005910.  
XX  
PR 26-FEB-2001; 2001US-0271646P.  
PR 27-FEB-2001; 2001US-0271840P.  
PR 28-FEB-2001; 2001US-0272404P.  
PR 28-FEB-2001; 2001US-0272405P.  
PR 28-FEB-2001; 2001US-0272410P.  
PR 28-FEB-2001; 2001US-0272414P.  
PR 02-MAR-2001; 2001US-0272787P.  
PR 02-MAR-2001; 2001US-0272922P.  
PR 02-MAR-2001; 2001US-0273048P.  
PR 02-MAR-2001; 2001US-0273300P.  
PR 16-MAR-2001; 2001US-0276401P.  
PR 20-MAR-2001; 2001US-0277324P.  
PR 20-MAR-2001; 2001US-0278660P.  
PR 30-MAR-2001; 2001US-0280039P.  
PR 30-MAR-2001; 2001US-0280234P.  
PR 02-APR-2001; 2001US-0280818P.  
PR 12-APR-2001; 2001US-0283443P.  
PR 23-APR-2001; 2001US-0285754P.  
PR 03-MAY-2001; 2001US-0286096P.  
PR 17-MAY-2001; 2001US-0288353P.  
PR 31-MAY-2001; 2001US-0291703P.  
PR 31-MAY-2001; 2001US-0294834P.  
PR 20-JUN-2001; 2001US-0299695P.  
PR 21-JUN-2001; 2001US-0299845P.  
PR 05-JUL-2001; 2001US-0303242P.  
PR 13-AUG-2001; 2001US-0311981P.  
PR 16-AUG-2001; 2001US-0312858P.  
PR 17-AUG-2001; 2001US-0313280P.  
PR 29-AUG-2001; 2001US-0315614P.  
PR 17-SEP-2001; 2001US-0322818P.  
PR 25-FEB-2002; 2002US-00322818.

XX (CURA-) CURAGEN CORP.  
XX  
XX Alsobrook JP, Anderson DW, Ballinger RA, Boldog FL, Burgess CE;  
PI Casman SJ, Ellerman KE, Gangolli EA, Gerlach VL, Gilbert JA;  
PI Gorman L, Guo X, Gusev VY, Kekuda R, Li L, Liu X, Malyankar UM;  
PI Miller CE, Millet I, Padigar M, Patturajan M, Pena CE, Peyman JA;  
PI Rastelli L, Shenoy SG, Shinkets RA, Smithson G, Spytek KA, Stone DJ;  
PI Taupier RJ, Tchernev VT, Vernet CAM, Zerhusen BD;  
XX

DR WPI: 2002-698672/75.  
XX N-PSDB; ADH48739.  
PT New NOVX polypeptides or polynucleotides, useful for preventing or  
PT treating disorders or syndromes e.g., atherosclerosis, hypertension,  
PT obesity or cancer.  
XX  
PS Claim 1; Page 74; 923pp; English.  
XX  
CC The present invention relates to novel human NOVX proteins, where X is  
CC any number from 1 to 91 and their coding sequences. The proteins and  
CC coding sequences are useful for preventing or treating disorders or  
CC syndromes e.g. atherosclerosis, hypertension, obesity or cancer. NOV10 is  
CC a T01C1.3-like protein and its coding sequence maps to chromosome 4.  
XX  
SQ Sequence 356 AA;  
Query Match 78.0%; Score 1754; DB 5; Length 356;  
Best Local Similarity 100.0%; Pred. No. 2.1e-134; Indels 0; Gaps 0;  
Matches 343; Conservative 0; Mismatches 0;  
QY 100 RLHNRREDYVERSAEFADGLLSKALKDIQSGALDINKAGILYGPQKTLHLLEALPAGK 159  
DB 14 RLHNRREDYVERSAEFADGLLSKALKDIQSGALDINKAGILYGPQKTLHLLEALPAGK 73  
QY 160 PASFNKTRDFHDSYSYKDSKETCAVLQKVALWARAQARTKSKLNLETSEIKPTAS 219  
DB 74 PASFNKTRDFHDSYSYKDSKETCAVLQKVALWARAQARTKSKLNLETSEIKPTAS 133  
QY 220 TYLHOLTLLQKMYTOPKEKNESIQYETSNPTVOLKIPOLRVSSVSKSQPDGSLDDVMYQV 279  
DB 134 TYLHOLTLLQKMYTOPKEKNESIQYETSNPTVOLKIPOLRVSSVSKSQPDGSLDDVMYQV 193  
QY 280 SKTSSVLEGSALQKLNILPKQNKIECSGPTVTHSSVDSYFLHGDLSPLCLNSKNGTVDGT 339  
DB 194 SKTSSVLEGSALQKLNILPKQNKIECSGPTVTHSSVDSYFLHGDLSPLCLNSKNGTVDGT 253  
QY 340 SENTEDGLDRKDSKQPRKGRYQYDHEIMEEAIAMVMSGKMSVSKAQGIYGVPHSTLE 399  
DB 254 SENTEDGLDRKDSKQPRKGRYQYDHEIMEEAIAMVMSGKMSVSKAQGIYGVPHSTLE 313  
QY 400 YKVKERSGTLKTPPKKKLRLPDTGLYNMTDSGTGCKNSKSPV 442  
DB 314 YKVKERSGTLKTPPKKKLRLPDTGLYNMTDSGTGCKNSKSPV 356  
RESULT 4  
ABO60261  
ID ABO60261 standard; protein; 277 AA.  
XX  
AC ABO60261;  
DT 29-JUL-2004 (first entry)  
XX  
DE Human genome derived single exon protein #6495.  
XX Human; gene expression; single exon probe; microarray;  
KW alternative splicing event; genomic alteration.  
XX  
OS Homo sapiens.  
XX  
PN US2003194704-A1.  
XX  
PD 16-OCT-2003.  
XX  
PF 03-APR-2002; 2002US-00029386.  
XX  
PR 03-APR-2002; 2002US-00029386.  
XX  
PA (PENN/) PENN S G.  
PA (RANK/) RANK D R.  
PA (HANZ/) HANZEL D K.  
XX

Thu Feb 24 10:11:18 2005

us-10-016-768a-8.rag

PI Penn SG, Rank DR, Hanzel DK;  
XX WPI; 2004-119264/12.  
XX New human genome-derived single exon nucleic acid probes useful for human  
PT gene expression analysis, for identifying or characterizing alternative  
PT splicing events, for assessing genomic alterations or as tools for  
PT surveying tissues.  
XX  
XX Claim 45; SEQ ID NO 33895; 80pp; English.  
PS  
XX The invention relates to a nucleic acid probe for measuring human gene  
CC expression, comprising any of the 27,400 fully defined nucleotide  
CC sequences in the specification, or their complements or fragments, and  
CC encoding at least 8 amino acids of any of the 6888 amino acid sequences  
CC fully defined in the specification. The probe is a single exon probe that  
CC hybridizes under high stringency conditions to a nucleic acid molecule  
CC expressed in human cells or tissues. Also included are a spatially-  
CC addressable set of single exon nucleic acid probes for measuring human  
CC gene expression (comprising a plurality of single exon nucleic acid  
CC probes cited above, where each of the plurality of probes is separately  
CC and addressably isolatable or amplifiable from the plurality), a single  
CC exon microarray for measuring human gene expression, a method of  
CC measuring human gene expression, a vector comprising the single exon  
CC probe cited above, an ORF-encoded peptide comprising at least 8  
CC contiguous amino acids of any of the above-mentioned amino acid  
CC sequences (optionally with conservative amino acid substitutions), an  
CC isolated antibody that binds specifically to a peptide cited above,  
CC methods of selling and/or licensing single exon probes or microarrays to  
CC a customer desiring to measure gene expression, a method of providing  
CC human gene expression data by subscription, and a computer-readable  
CC storage medium which contains a database having a plurality of records  
CC (each record including data on the expression of a single exon probe  
CC cited above. The probe, methods and apparatus are useful in gene  
CC expression analysis. The probes may be used as tools for surveying  
CC tissues to detect the presence of expressed messages that contain their  
CC specific exon, or in constructing genome-derived single exon microarrays.  
CC In addition, the probes are used in identifying and characterizing  
CC alternative splicing events, in detecting and characterizing gross  
CC alterations in the genomic locus that includes their exon, in assessing  
CC smaller genomic alterations, in priming the synthesis of nucleic acids,  
CC or in expressing the ORF-encoded peptide. The present sequence is a human  
CC single exon probe protein of the invention. Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=20030194704  
XX  
SQ Sequence 277 AA;  
Query Match 62.7%; Score 1411; DB 8; Length 277;  
Best Local Similarity 100.0%; Pred. No. 1.4e-106;  
Matches 277; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 100 RLHRNREYVERSAEADGALLSKALDIOQALDINKAGILYGIPOKTLHLLEALPAGK 159  
DB 1 RLHRNREYVERSAEADGALLSKALDIOQALDINKAGILYGIPOKTLHLLEALPAGK 60  
QY 160 PASFNKTRPHDSYSYKDSKETCAVLQKVALWARAQARTKSKLNLETSEIKFPTAS 219  
DB 61 PASFNKTRPHDSYSYKDSKETCAVLQKVALWARAQARTKSKLNLETSEIKFPTAS 120  
QY 220 TYLHQLTLQKMWVTFQKKNESLOYETSNPTVOLKIPOLRVSSVSKSQPDGSLLDVMYQV 279  
DB 121 TYLHQLTLQKMWVTFQKKNESLOYETSNPTVOLKIPOLRVSSVSKSQPDGSLLDVMYQV 180  
QY 280 SKTSVLEGSALQKLNILPKONKIECSGPVTHSSVDSYFLHGLDLSPLCLNSKNGTVDTGT 339  
DB 181 SKTSVLEGSALQKLNILPKONKIECSGPVTHSSVDSYFLHGLDLSPLCLNSKNGTVDTGT 240  
QY 340 SENTEDGLDRKDSKQPKRGRYRQYDHEIMEEAIAM 376  
DB 241 SENTEDGLDRKDSKQPKRGRYRQYDHEIMEEAIAM 277

RESULT 5  
ADA54381  
ID ADA54381 standard; protein; 213 AA.  
XX  
AC ADA54381;  
XX  
DT 20-NOV-2003 (first entry)  
XX  
DE Human protein, SEQ ID 1949.  
XX  
KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;  
KW Gene therapy; human; secretory protein; membrane proteins; cancer;  
KW inflammatory disease; osteoporosis; neurological disease.  
OS Homo sapiens.  
XX EPI293569-A2.  
XX  
PD 19-MAR-2003.  
XX  
PF 21-MAR-2002; 2002EP-00006586.  
XX  
PR 14-SEP-2001; 2001JP-00328381.  
PR 24-JAN-2002; 2002US-0350435P.  
XX  
PA (HELI-) HELIX RES INST.  
PA (REAS-) RES ASSOC BIOTECHNOLOGY.  
XX  
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Matsuho Y;  
XX  
DR WPI; 2003-395539/38.  
DR N-PSDB; ADA52742.  
XX  
PT New polynucleotides encoding full-length polypeptides, e.g. secretory  
PT and/or membrane proteins, useful for developing medicines for diseases in  
PT which the gene is involved, or as target molecules for gene therapy.  
XX  
PS Claim 14; SEQ ID NO 1949; 205pp; English.  
XX  
CC The present invention relates to novel human secretory or membrane  
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-  
CC ADA54071). The coding sequences are useful in the gene therapy of  
CC diseases caused by abnormalities of the proteins, e.g. cancer,  
CC inflammatory diseases, osteoporosis or neurological disease.  
XX  
SQ Sequence 213 AA;  
Query Match 48.4%; Score 1090; DB 6; Length 213;  
Best Local Similarity 99.5%; Pred. No. 1.5e-80;  
Matches 212; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 230 WYTFQKKNESLOYETSNPTVOLKIPOLRVSSVSKSQPDGSLLDVMYQVSKTSVLEGS 289  
DB 1 WYTFQKKNESLOYETSNPTVOLKIPOLRVSSVSKSQPDGSLLDVMYQVSKTSVLEGS 60  
QY 290 ALQKLNILPKONKIECSGPVTHSSVDSYFLHGLDLSPLCLNSKNGTVDTGTSENTEDGLDR 349  
DB 61 ALQKLNILPKONKIECSGPVTHSSVDSYFLHGLDLSPLCLNSKNGTVDTGTSENTEDGLDR 120  
QY 350 KDSKQPKRGRYRQYDHEIMEEAIAMVWSGKMSVSKAQGIYGVPHSTILEYVKERSGTL 409  
DB 121 KDSKQPKRGRYRQYDHEIMEEAIAMVWSGKMSVSKAQGIYGVPHSTILEYVKERSGTL 180  
QY 410 KTFPPKKLRLPDTGLYNMTDSTGTCCKNSKXPV 442  
DB 181 KTFPPKKLRLPDTGLYNMTDSTGTCCKNSKXPV 213  
RESULT 6  
ADM04876

```

ID  ADM04876 standard; protein; 311 AA.
XX
AC  ADM04876;
XX
DT  20-MAY-2004 (first entry)
XX
DE  Human protein of the invention SEQ ID NO:3561.
XX
KW  human; gene therapy; diagnostic marker; pharmaceutical.
XX
OS  Homo sapiens.
XX
PN  EP1347046-A1.
XX
PD  24-SEP-2003.
XX
PF  12-APR-2002; 2002EP-00008400.
XX
PR  22-MAR-2002; 2002JP-00137785.
XX
PA  (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI  Isoqai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI  Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI  Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR  WPI; 2003-723558/69.
DR  N-PSDB; ADM02433.
XX
XX  New polynucleotides and polypeptides are useful in gene therapy, for
PT  developing a diagnostic marker or medicines for regulating their
PT  expression and activity, or as a target of gene therapy.
XX
PS  Claim 1; SEQ ID NO 3561; 305pp; English.
XX
CC  The invention relates to a novel human polynucleotide and the encoded
CC  polypeptide. A polynucleotide of the invention may have a use in gene
CC  therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful
CC  as a primer for synthesizing the polynucleotide or as a probe for
CC  detecting the polynucleotide. The polynucleotides ADM01316-ADM03758 are
CC  useful in gene therapy, for developing a diagnostic marker or medicines
CC  for regulating their expression and activity, or as a target of gene
CC  therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides
CC  are useful as pharmaceutical agents. The present sequence represents a
CC  protein sequence of the invention.
XX
SQ  Sequence 311 AA;

Query Match 22.3%; Score 502; DB 7; Length 311;
Best Local Similarity 89.5%; Pred. No. 2.5e-32;
Matches 102; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY  1 MKKMIQFAIEYISKGKTOENRNSIGPSIVCKSIQNMNAENSLQEEQGLDLTVNRM 60
DB  154 MKKMIQFAIEYISKGKTOENRNSIGPSIVCKSIQNMNAENSLQEEQGLDLTVNRM 213
QY  61 QSONTQQGGVLDLSTKTKTSIKSESSICDPSSSENSVAGRLNRNEDYVERSAE 114
DB  214 QSONTQQGGVLDLSTKTKTSIKSESSICDPSSSENSVAGRLNRNEDYVERSAE 267

RESULT 7
ADN35086
ID  ADN35086 standard; protein; 433 AA.
XX
AC  ADN35086;
XX
DT  18-NOV-2004 (first entry)
XX
DE  Nuclear receptor transcriptional corepressor.
XX
KW  co-repressor; nuclear receptor; ligand-dependent transactivation;
KW  steroid hormone signal.
XX
ADN04876 standard; protein; 311 AA.
XX
AC  ADM04876;
XX
DT  20-MAY-2004 (first entry)
XX
DE  Human protein of the invention SEQ ID NO:3561.
XX
KW  human; gene therapy; diagnostic marker; pharmaceutical.
XX
OS  Homo sapiens.
XX
PN  EP1347046-A1.
XX
PD  24-SEP-2003.
XX
PF  12-APR-2002; 2002EP-00008400.
XX
PR  22-MAR-2002; 2002JP-00137785.
XX
PA  (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI  Isoqai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI  Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI  Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR  WPI; 2003-723558/69.
DR  N-PSDB; ADM02433.
XX
XX  New polynucleotides and polypeptides are useful in gene therapy, for
PT  developing a diagnostic marker or medicines for regulating their
PT  expression and activity, or as a target of gene therapy.
XX
PS  Claim 1; SEQ ID NO 3561; 305pp; English.
XX
CC  The invention relates to a novel human polynucleotide and the encoded
CC  polypeptide. A polynucleotide of the invention may have a use in gene
CC  therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful
CC  as a primer for synthesizing the polynucleotide or as a probe for
CC  detecting the polynucleotide. The polynucleotides ADM01316-ADM03758 are
CC  useful in gene therapy, for developing a diagnostic marker or medicines
CC  for regulating their expression and activity, or as a target of gene
CC  therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides
CC  are useful as pharmaceutical agents. The present sequence represents a
CC  protein sequence of the invention.
XX
SQ  Sequence 311 AA;

Query Match 22.3%; Score 502; DB 7; Length 311;
Best Local Similarity 89.5%; Pred. No. 2.5e-32;
Matches 102; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY  1 MKKMIQFAIEYISKGKTOENRNSIGPSIVCKSIQNMNAENSLQEEQGLDLTVNRM 60
DB  154 MKKMIQFAIEYISKGKTOENRNSIGPSIVCKSIQNMNAENSLQEEQGLDLTVNRM 213
QY  61 QSONTQQGGVLDLSTKTKTSIKSESSICDPSSSENSVAGRLNRNEDYVERSAE 114
DB  214 QSONTQQGGVLDLSTKTKTSIKSESSICDPSSSENSVAGRLNRNEDYVERSAE 267

RESULT 7
ADN35086
ID  ADN35086 standard; protein; 433 AA.
XX
AC  ADN35086;
XX
DT  18-NOV-2004 (first entry)
XX
DE  Nuclear receptor transcriptional corepressor.
XX
KW  co-repressor; nuclear receptor; ligand-dependent transactivation;
KW  steroid hormone signal.
XX
ADN04876 standard; protein; 311 AA.
XX
AC  ADM04876;
XX
DT  20-MAY-2004 (first entry)
XX
DE  Human protein of the invention SEQ ID NO:3561.
XX
KW  human; gene therapy; diagnostic marker; pharmaceutical.
XX
OS  Homo sapiens.
XX
PN  EP1347046-A1.
XX
PD  24-SEP-2003.
XX
PF  12-APR-2002; 2002EP-00008400.
XX
PR  22-MAR-2002; 2002JP-00137785.
XX
PA  (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI  Isoqai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI  Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI  Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR  WPI; 2003-723558/69.
DR  N-PSDB; ADM02433.
XX
XX  New polynucleotides and polypeptides are useful in gene therapy, for
PT  developing a diagnostic marker or medicines for regulating their
PT  expression and activity, or as a target of gene therapy.
XX
PS  Claim 1; SEQ ID NO 3561; 305pp; English.
XX
CC  The invention relates to a novel human polynucleotide and the encoded
CC  polypeptide. A polynucleotide of the invention may have a use in gene
CC  therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful
CC  as a primer for synthesizing the polynucleotide or as a probe for
CC  detecting the polynucleotide. The polynucleotides ADM01316-ADM03758 are
CC  useful in gene therapy, for developing a diagnostic marker or medicines
CC  for regulating their expression and activity, or as a target of gene
CC  therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides
CC  are useful as pharmaceutical agents. The present sequence represents a
CC  protein sequence of the invention.
XX
SQ  Sequence 311 AA;

Query Match 22.2%; Score 499.5; DB 8; Length 433;
Best Local Similarity 33.6%; Pred. No. 6.3e-32;
Matches 158; Conservative 66; Mismatches 143; Indels 103; Gaps 20;

QY  1 MKKMIQFAIEYISKGKTOE-----NRNGS-----IGPSIVCKSIQNMNAENSLQEE 48
DB  1 MORMIQQAFAEYTSKNSSTQDPNQSNSTKNQSLKASPVTTSTTAATTQNVLSKLLMAD 60
QY  49 QSGPLDLTVNRMQEQNTQGGVLDLSTKTKT-----SIKSESSICDPSSSENSVAGRLNRN 104
DB  61 QSGPLDLTVNRMQEQNTQGGVLDLSTKTKT-----SIKSESSICDPSSSENSVAGRLNRN 119
QY  105 REDYVERSAEAFADGLLSKALKD-----IQSGALDINKAGILYIPQKTLHLHLALPAG 158
DB  120 RPD-----GLRSGDGVPPRSLOQDGTREGFCHSTSLKVPILA-----RSLQISELLSRN 167
QY  159 K-----PASFKNTRDFHDSYSYKOSKETCAVLQKVALWARAQAE-RTEKSKLN--- 206
DB  158 QLSTAASLPGSLQN-----HGQH-----LILSREASWAKPHYENLGRMKFRNG 213
QY  207 -LLETSEIKFPTASTYLHQLTQKMTQFKENESLOYETSNTPTVQLKIPQRLRVSSVSXS 265
DB  214 ALSNISDLPFLAENSAFFKMLQ-----AKQGGKGVSHSSPVDLKIPIQVRGMDLSWE 266
QY  266 QPDGSGLLDMVYQVSKTSVL-----EGSALQKLNILPKQNKIEC--SGPVTHSSVDSYF 319
DB  267 SRTGD-----QYSYSSILVWGSQTESALSCKLRAILPKQSRKSMLDAGP-----DSWG 313
QY  320 LHGDLSPCLNSKNGTVDGTSNTEDGLDRKDSKQPRKRGYRQYDHIEMEAIVMS 379
DB  314 SDAE-----QSTSGQPYPTSDQEGD-----PGSKQPRKRGYRQYNSILBEAISVMS 363
QY  380 GKMSVSKAQGIYGVPHSTLEYKVKERSGTLTTPPKKLLRL-----PDTCL 424
DB  364 GKMSVSKAQSIYGIPIHSTLEYKVKERLGTLPKPPKKMKLMRSEGPDSV 413

```





KW myeloma; neutropenic; vasotropic; immunostimulant; cerebroprotective;  
 KW cardiant; E93 protein.  
 OS Homo sapiens.  
 XX WO200234882-A2.  
 XX 02-MAY-2002.  
 XX 29-OCT-2001; 2001WO-US048053.  
 XX 27-OCT-2000; 2000US-0243865P.  
 XX (UYMA-) UNIV MARYLAND BIOTECHNOLOGY INST.  
 XX Baehrecke EH;  
 XX WPI; 2002-479717/51.  
 XX Novel programmed cell death modulating proteins, useful for treating or  
 PT preventing disorders associated with abnormal cell proliferation and  
 PT apoptosis such as cancer, stroke, Parkinson's disease, myocardial  
 PT infarction.  
 XX Claim 1; Fig 1; 88pp; English.

KW vaccine; gene therapy; lymphoma; cytostatic; melanoma; neuroprotective;  
 KW myeloma; neutropenic; vasotropic; immunostimulant; cerebroprotective;  
 KW cardiant; E93 protein.  
 OS Tetraodon nigroviridis.  
 XX WO200234882-A2.  
 XX 02-MAY-2002.  
 XX 29-OCT-2001; 2001WO-US048053.  
 XX 27-OCT-2000; 2000US-0243865P.  
 XX (UYMA-) UNIV MARYLAND BIOTECHNOLOGY INST.  
 XX Baehrecke EH;  
 XX WPI; 2002-479717/51.  
 XX Novel programmed cell death modulating proteins, useful for treating or  
 PT preventing disorders associated with abnormal cell proliferation and  
 PT apoptosis such as cancer, stroke, Parkinson's disease, myocardial  
 PT infarction.  
 XX Claim 1; Fig 1; 88pp; English.

KW myeloma; neutropenic; vasotropic; immunostimulant; cerebroprotective;  
 KW cardiant; E93 protein.  
 OS Homo sapiens.  
 XX WO200234882-A2.  
 XX 02-MAY-2002.  
 XX 29-OCT-2001; 2001WO-US048053.  
 XX 27-OCT-2000; 2000US-0243865P.  
 XX (UYMA-) UNIV MARYLAND BIOTECHNOLOGY INST.  
 XX Baehrecke EH;  
 XX WPI; 2002-479717/51.  
 XX Novel programmed cell death modulating proteins, useful for treating or  
 PT preventing disorders associated with abnormal cell proliferation and  
 PT apoptosis such as cancer, stroke, Parkinson's disease, myocardial  
 PT infarction.  
 XX Claim 1; Fig 1; 88pp; English.

The present invention relates to novel programmed cell death modulating  
 CC proteins and polynucleotides encoding such proteins. Sequences of the  
 CC invention are useful to screen potential cellular apoptosis inhibiting  
 CC compounds to determine their use as therapeutic agents for treatment of  
 CC diseases associated with increased programmed cell death. They are also  
 CC useful for treating or preventing disorders associated with decrease in  
 CC apoptosis. Programmed cell death modulating sequences are useful for  
 CC treating or preventing cancer e.g. adenocarcinoma, leukaemia, lymphoma,  
 CC melanoma, myeloma. Inhibition of the activity of the sequences of the  
 CC invention are useful for treating disorders associated with increase in  
 CC cell death or apoptosis such as acquired immunodeficiency syndrome  
 CC (AIDS), neurodegenerative diseases (e.g., Alzheimer's disease, retinitis  
 CC pigmentosa, Parkinson's disease and cerebellar degeneration), ischaemic  
 CC injuries (e.g., myocardial infarction, stroke, reperfusion injury),  
 CC myelodysplastic syndromes (e.g., aplastic anaemia), toxin-induced  
 CC diseases and other infectious or genetic immunodeficiencies. Sequences of  
 CC the invention are used as vaccines and in gene therapy. The present  
 CC sequence is human E93 programmed cell death modulating protein conserved  
 CC domain  
 XX Sequence 53 AA;  
 SQ

Query Match 12.1%; Score 273; DB 5; Length 53;  
 Best Local Similarity 100.0%; Pred. No. 9.4e-15;  
 Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 353 KQPRKKRGYQYDHEIMEEAIAMVMSGKMSVSKAQGIYGVPHSTLEYKVKER 405  
 DB 1 KQPRKKRGYQYDHEIMEEAIAMVMSGKMSVSKAQGIYGVPHSTLEYKVKER 53

RESULT 12  
 AAE24593  
 ID AAE24593 standard; protein; 54 AA.  
 XX  
 AC AAE24593;  
 XX  
 DT 04-OCT-2002 (first entry)  
 XX  
 DE Fish E93 programmed cell death modulating protein conserved domain.

Fish; cancer; programmed cell death modulating protein; adenocarcinoma;  
 KW cellular apoptosis; leukaemia; acquired immunodeficiency syndrome; AIDS;  
 KW neurodegenerative disease; Alzheimer's disease; retinitis pigmentosa;  
 KW Parkinson's disease; myelodysplastic syndrome; cerebellar degeneration;  
 KW aplastic anaemia; ischaemic injury; myocardial infarction; stroke;  
 KW reperfusion injury; toxin-induced disease; genetic immunodeficiency;

Sequence 54 AA;

Query Match 10.4%; Score 233.5; DB 5; Length 54;  
 Best Local Similarity 81.5%; Pred. No. 1.6e-11;  
 Matches 44; Conservative 7; Mismatches 2; Indels 1; Gaps 1;

QY 353 KQPRKKRGYQYDHEIMEEAIAMVMSGKMSVSKAQGIYGVPHSTLEYKVKER 405  
 DB 1 KQPRKKRGYQYDHEIMEEAIAMVMSGKMSVSKAQGIYGVPHSTLEYKVKER 54

RESULT 13  
 AAE24594  
 ID AAE24594 standard; protein; 53 AA.  
 XX  
 AC AAE24594;  
 XX  
 DT 04-OCT-2002 (first entry)  
 XX  
 DE Mouse E93 programmed cell death modulating protein conserved domain.

Mouse; cancer; programmed cell death modulating protein; adenocarcinoma;  
 KW cellular apoptosis; leukaemia; acquired immunodeficiency syndrome; AIDS;  
 KW neurodegenerative disease; Alzheimer's disease; retinitis pigmentosa;  
 KW Parkinson's disease; myelodysplastic syndrome; cerebellar degeneration;  
 KW aplastic anaemia; ischaemic injury; myocardial infarction; stroke;



KW reperfusion injury; toxin-induced disease; genetic immunodeficiency;  
KW vaccine; gene therapy; lymphoma; cytostatic; melanoma; neuroprotective;  
KW myeloma; nontropic; vasotropic; immunostimulant; cerebroprotective;  
KW cardiant; E93 protein.  
XX  
OS Mus musculus.  
XX  
PN WO200234882-A2.  
XX  
PD 02-MAY-2002.  
XX  
XX 29-OCT-2001; 2001WO-US048053.  
XX  
XX 27-OCT-2000; 2000US-0243865P.  
PR  
XX (UYMA-) UNIV MARYLAND BIOTECHNOLOGY INST.  
PA  
XX Baehrecke EH;  
XX  
XX WPI; 2002-479717/51.  
DR  
XX Novel programmed cell death modulating proteins, useful for treating or  
PT preventing disorders associated with abnormal cell proliferation and  
PT apoptosis such as cancer, stroke, Parkinson's disease, myocardial  
PT infarction.  
XX  
PS Claim 1; Fig 1; 88pp; English.  
XX  
CC The present invention relates to novel programmed cell death modulating  
CC proteins and polynucleotides encoding such proteins. Sequences of the  
CC invention are useful to screen potential cellular apoptosis inhibiting  
CC compounds to determine their use as therapeutic agents for treatment of  
CC diseases associated with increased programmed cell death. They are also  
CC useful for treating or preventing disorders associated with decrease in  
CC apoptosis. Programmed cell death modulating sequences are useful for  
CC treating or preventing cancer e.g. adenocarcinoma, leukaemia, lymphoma,  
CC melanoma, myeloma. Inhibition of the activity of the sequences of the  
CC invention are useful for treating disorders associated with increase in  
CC cell death or apoptosis such as acquired immunodeficiency syndrome  
CC (AIDS), neurodegenerative diseases (e.g., Alzheimer's disease, retinitis  
CC pigmentosa, Parkinson's disease and cerebellar degeneration), ischaemic  
CC injuries (e.g., myocardial infarction, stroke, reperfusion injury),  
CC myelodysplastic syndromes (e.g., aplastic anaemia), toxin-induced  
CC diseases and other infectious or genetic immunodeficiencies. Sequences of  
CC the invention are used as vaccines and in gene therapy. The present  
CC sequence is mouse E93 programmed cell death modulating protein conserved  
CC domain  
XX  
SQ Sequence 53 AA;  
Query Match 10.2%; Score 229; DB 5; Length 53;  
Best Local Similarity 81.1%; Pred. No. 3.6e-11;  
Matches 43; Conservative 6; Mismatches 4; Indels 0; Gaps 0;  
QY 353 KQPRKRGYQYDHEIMEEALAMVSGKMSVSKAGIYGVPHSTLEYKVKR 405  
Db 1 KHPRKRGYQYNSILEPISVLMSGKMSVSKAGIYGVPHSTLEYKVKR 53  
RESULT 14  
ABB71145  
ID ABB71145 standard; protein; 1140 AA.  
XX  
XX ABB71145;  
AC  
XX 26-MAR-2002 (first entry)  
DT  
XX Drosophila melanogaster polypeptide SEQ ID NO 40227.  
DE  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.

XX WO200171042-A2.  
PN  
XX 27-SEP-2001.  
PD  
XX 23-MAR-2001; 2001WO-US009231.  
PF  
XX 23-MAR-2000; 2000US-0191637P.  
PR  
XX 11-JUL-2000; 2000US-00614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI  
XX WPI; 2001-656860/75.  
DR  
XX N-PSDB; ABL15248.  
DR  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions.  
XX  
PS Disclosure; SEQ ID NO 40227; 2ipp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB2072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1140 AA;  
Query Match 8.9%; Score 200.5; DB 4; Length 1140;  
Best Local Similarity 21.7%; Pred. No. 6.2e-07;  
Matches 92; Conservative 66; Mismatches 146; Indels 119; Gaps 14;  
QY 19 TCENRNGSIGPSIVCKSIQMNQAEISLQEQSGPLDITVNRMOEQNTQDGDVLDL--ST 76  
Db 497 SOENENGNASLLQQQHQHQHQQHQQHQQHQQHQQHQQHQQHQQHQQHQQHQQHQQH 556  
QY 77 KXTSKSEESSICDPSENSVAGRLHRNREDYVERSAEFADGLLSKALDKIQSGALDINK 136  
Db 557 EDPILKIPSKVSGPASSSS-----LSP 579  
QY 137 AGILYIPIOKTLLHLEALPAGKPASFKNTRDFHDSYK-----SKETCAVLQKVALW 192  
Db 580 GGLVGG-----HHHPLNNNSLSISNNSN--HSSNSHRNGSNRPHSASPLAAAV-- 628  
QY 193 ARAQAERTKSKLNLETSEIKFPTASTYHLQITLQKMYTQFKENKESLQYETSNFTVL 252  
Db 629 --AQQGYGAGNSLLTSSSSSIQMMASNIQRI-----NEQSGQES----- 667  
QY 253 KIPQLRVSSVSKSQPDGSGLLDMYQ-----VSKTSSVLEGSALQKLNLPKQNKIECS 307  
Db 668 ----LRNGNVSDCSNNGSGSSSLGYKKPSISVAKIIGGTDTSRFGASPNLLSQH----- 718  
QY 308 GPVTHSSVDSPYFLHGDLSPLCLNSKNGTVGDTSENTEGDLRKDS--KQPRKRGYQY 365  
Db 719 ----HS---AHHL-----THQQQQQQLSAQALGKGTGPRKRGYRNY 753  
QY 366 DHEIMEEALAMVSGKMSVSKAGIYGVPHSTLEYKVKERSGTLKTPPKKLRLLPTGLY 425  
Db 754 DRDSLVEAVKAVQRGEMSVHRAGSYGVPHSTLEYKVKERH---LMRPRKRPKPPQDLV 810  
QY 426 NMT 428  
Db 811 GLT 813



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 24, 2005, 01:45:39 ; Search time 44 Seconds  
(without alignments)  
749.884 Million cell updates/sec

Title: US-10-016-768A-8

Perfect score: 2250

Sequence: 1 MKKMTROFAIEYISKGTQ.....GLYNTDSTGSGCKNSKRPV 442

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/prodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/prodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/prodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/prodata/1/iaa/PCITUS\_COMB.pep.\*  
6: /cgn2\_6/prodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	131.5	5.8	2375	4	US-09-538-092-1131
2	130	5.8	3248	1	Sequence 1131, Ap
3	130	5.8	3248	5	Sequence 1, Appli
4	127	5.6	534	4	PCT-US95-16216-1
5	127	5.6	1753	4	Sequence 184, App
6	126.5	5.6	750	4	US-09-538-092-184
7	124.5	5.5	2954	4	Sequence 19154, A
8	123	5.5	2482	1	Sequence 3544, Ap
9	123	5.5	3210	4	Sequence 1, Appli
10	122	5.4	1164	4	Sequence 6, Appli
11	122	5.4	1938	4	Sequence 1154, Ap
12	122	5.4	1959	4	Sequence 399, App
13	122	5.4	2781	4	Sequence 6417, Ap
14	122	5.4	2907	4	Sequence 8134, Ap
15	119.5	5.3	475	4	Sequence 10, Appli
16	118	5.2	924	4	Sequence 1, Appli
17	118	5.2	1147	4	Sequence 14568, A
18	117	5.2	1146	3	Sequence 18798, A
19	117	5.2	1940	4	Sequence 1074, Ap
20	117	5.2	1963	4	Sequence 6, Appli
21	115.5	5.1	868	4	Sequence 901, App
22	114.5	5.1	569	4	Sequence 8886, Ap
23	114.5	5.1	1279	3	Sequence 9731, Ap
24	114.5	5.1	1279	4	Sequence 18, Appli
25	114.5	5.1	1279	4	Sequence 2, Appli
26	114.5	5.1	1279	4	Sequence 2, Appli
27	114.5	5.1	1786	4	Sequence 7880, Ap
28	114.5	5.1	2476	4	Sequence 7, Appli

28	114	5.1	1260	4	US-09-248-796A-20258	Sequence 20258, A
29	113.5	5.0	431	4	US-09-286-981B-3	Sequence 3, Appli
30	113.5	5.0	905	4	US-09-248-796A-16333	Sequence 16333, A
31	113	5.0	593	4	US-09-270-767-41549	Sequence 41549, A
32	113	5.0	876	4	US-09-764-176-10	Sequence 10, Appli
33	113	5.0	1531	4	US-09-949-016-6740	Sequence 6740, Ap
34	112.5	5.0	1444	4	US-09-949-016-9652	Sequence 9652, Ap
35	112.5	5.0	2004	4	US-09-538-092-1371	Sequence 1371, Ap
36	112.5	5.0	2004	4	US-09-949-016-6756	Sequence 6756, Ap
37	111.5	5.0	1564	4	US-10-144-198-2	Sequence 2, Appli
38	111.5	5.0	1564	4	US-10-144-198-4	Sequence 4, Appli
39	111.5	5.0	1780	4	US-09-949-016-6899	Sequence 6899, Ap
40	111.5	5.0	1972	4	US-08-875-435B-3	Sequence 3, Appli
41	111	4.9	1589	3	US-08-755-587-189	Sequence 189, App
42	111	4.9	1786	3	US-08-973-462-8	Sequence 8, Appli
43	111	4.9	2753	4	US-09-949-016-7659	Sequence 7659, Ap
44	111	4.9	2753	4	US-09-949-016-7660	Sequence 7660, Ap
45	111	4.9	3924	4	US-09-538-092-1246	Sequence 1246, Ap

#### ALIGNMENTS

RESULT 1

US-09-538-092-1131

; Sequence 1131, Application US/09538092

; Patent No. 6753314

; GENERAL INFORMATION:

; APPLICANT: Giot, Loic

; APPLICANT: Mansfield, Traci A.

; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same

; FILE REFERENCE: 15966-542

; CURRENT APPLICATION NUMBER: US/09/538,092

; CURRENT FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: 60/127,352

; PRIOR FILING DATE: 1999-04-01

; PRIOR APPLICATION NUMBER: 60/178,965

; PRIOR FILING DATE: 2000-02-01

; NUMBER OF SEQ ID NOS: 1387

; SOFTWARE: CurapatSeqFormat Version 0.9

; SEQ ID NO 1131

; LENGTH: 2375

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (0)..(0)

; OTHER INFORMATION: Polypeptide Accession Number P46100

US-09-538-092-1131

Query Match 5.8%; Score 131.5; DB 4; Length 2375;  
Best Local Similarity 23.0%; Pred. No. 0.02;  
Matches 100; Conservative 63; Mismatches 173; Indels 99; Gaps 21;

QY 14 SKSGKTQENRNG-STGPSIVCKSIQMNQAEISLOEQEGPLDLTVNMQEONTQQDGV 72

Db 667 STSGSDFDTKKGSAKSIISKKRTQSESS---NYDSELEKEIKSMKIGAAR----- 718

QY 73 DLSTTK--TSIKSESSICDPSSNSVAGRLHNRDYYVERSAEPADGLLSKALKDIQSG 130

Db 719 --TTKRIPTWTKDFPSSSEDEKSKGMDNQGHKLNKTSQEGSSDDAERKQERTFFSAEG 776

QY 131 ALDINKAGILYGIPTKTLHLH-EALPAGKPFASFKNKTDFHDSYSYKDSKETCAVLQKV 189

Db 777 TVD-----KDTTLMELRDLPLPKQAS---ASTDGVDKLKGKQSFTSLVRKV 822

QY 190 ALWARAQAERTSKKLNLLTSEIPEPTASTYLHQLTQKMTQPFKEKNEISLOYETSNPT 249

Db 823 -----AETKEKSK-----HLKTKTKKV--QDGLSDIAEKFLKQDS--DETSED 864

QY 250 VQLKIPQLRVSSVSKSQPDGSLDVMYOVSKTSSVLEGSALQKLNILPKQNKIESGSP 309

Db 865 KK-----QSKKGTEEEKKPS-----DFKKVKVIMQEQYSSSDGTEK--LPEREI-CHFP 912

Thu Feb 24 10:11:19 2005

QY 310 VTHSSVDSYFLHGLDPLCLNSKNGTVDTGTSNTEDGLDRKDSKQPRKRGYRQYDHEI 369  
DB 913 KGIKOI-----KNGTTIDG-----EKSKKIRDKTSKKKDELSDY 946  
QY 370 MEATAMVSGKMSVSK--AQGIYGVPHSTLBYKVKERSGTLKTPPKK--LRLPDTGLY 425  
DB 947 ABKSTGKSDSCSDSSDKSKNGAYG-----REKRCCKLLGKSSRKQDCSSDTEKY 998  
QY 426 NMTDSTGSGCKNSK 440  
DB 999 SMKEDG---CNSSDK 1010  
RESULT 2  
US-08-353-700-1  
; Sequence 1, Application US/08353700  
; Patent No. 5599919  
; GENERAL INFORMATION:  
; APPLICANT: YEN, TIMOTHY J.  
; APPLICANT: RATTNER, JEROME B.  
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A  
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,  
; TITLE OF INVENTION: AND METHODS OF USE  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN  
; STREET: 1601 MARKET STREET, SUITE 720  
; CITY: PHILADELPHIA  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103-2307  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/353,700  
; FILING DATE: 09-DEC-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: REED, JANET E.  
; REGISTRATION NUMBER: 36,252  
; TELEPHONE: (215) 563-4100  
; TELEFAX: (215) 563-4044  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3248 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: HUMAN  
US-08-353-700-1  
Query Match 5.8%; Score 130; DB 1; Length 3248;  
Best Local Similarity 20.2%; Pred. No. 0.045;  
Matches 104; Conservative 74; Mismatches 200; Indels 138; Gaps 21;  
QY 7 QFAIEYISGKTOENR---NGSIGPISVCKSIOMNQANSIQEQQEGPLDLTVNRMQEQ 63  
DB 2676 QDTLEVLOSQSYKNLENELELTQWDKMSFVEKVNKTAKETELQREHWEAQTAEQLQEL 2735  
QY 64 NTQQGQGVLDLSTKTKTSIKSESSICDPSSENS-VAGRLHNRREDYVERSFAFDGL--- 119  
DB 2736 SGEKNRFLAGLQLLLEIKSKDKQLKELTLENSELKSLDCMHKQDQVEKGVREIAEY 2795  
QY 120 ---LSKALKDIQSGALDINKAGILYIPQKTLILLHLEALPAGKPA5FKNKTDPHDSYSY 176

DB 2796 QLRHAEAKKHQALLDITNKQ---YEVEIQT---YREKL----- 2828  
QY 177 KDSKEICAVLOKVALWARAQARTEKSLN--LLETSEIKFPTASTYHLHQLTLQKMTQF 234  
DB 2829 -TSKECLSSQKLEI-----DLLKSSKEELNNSLKATQIILEELKTKMDNL---KYVNQL 2880  
QY 235 KEKNESL-----QYETSNPTVQKIPQLRVSSVSKSQPDGSLLDVWYQVSKT 282  
DB 2881 KKENERAQGRMKLLIKSKQLEBEKEILOKELSQLOAAQ-----EKQKT 2924  
QY 283 SSVLE-----GSALQKLNILPKQKIE-----CSGPVTHSSVD--SYFLHGLDLSPL 327  
DB 2925 GTVMDTKVDELTTTEIKELTLEKTEKPADEYLDKVCSSLISHEKLEKAKEMLETQVAHL 2984  
QY 328 C-----LNSKNGTVDG-----TSNTEDGLDRKDSKQPRK-----RGRYRQ 364  
DB 2985 CSQOSKQDSRGSPLPGVPVPGSPIPSVTEKRLSSGQNKASKGRQSSGIWENGRTTPA 3044  
QY 365 YDHEIMEEAIAMVSGKMSVSKAQGI-----YGVPH-----STLEYKVKERSGT 408  
DB 3045 TPESFSKSKKAVNGSIHPAEDTEGTEFEPEGLPEVVYKGFADIPITGKTSPIILRTTMA 3104  
QY 409 LKTPPK---KKRLPDTGLYNMTDSTGSGCKNSKP 441  
DB 3105 TRTSPRLAAQKALSLSL-----GKENLAESSKP 3134  
RESULT 3  
PCT-US95-16216-1  
; Sequence 1, Application PC/TUS9516216  
; GENERAL INFORMATION:  
; APPLICANT: Yen, Timothy J.  
; APPLICANT: Rattner, Jerome B.  
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently  
; TITLE OF INVENTION: Expressed Kinetochore Protein, and Methods of Use  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman  
; STREET: 1601 Market Street Suite 720  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103-2307  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/16216  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/353,700  
; FILING DATE: 09-DEC-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Reed, Janet E.  
; REGISTRATION NUMBER: 36,252  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (215) 563-4100  
; TELEFAX: (215) 563-4044  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3248 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
PCT-US95-16216-1

Query Match	5.8%;	Score 130;	DB 5;	Length 3248;
Best Local Similarity	20.2%;	Pred. No. 0.045;		
Matches 104;	Conservative 74;	Mismatches 200;	Indels 138;	Gaps 21;

  

QY	7	QFALEYISKSKTQENR---	NGSIGPSTVCKSIQMNQAENSLQESGQLDLTVNRMQEQ	63
DB	2676	QDTLEVQSSYKNULENELELTKMDKMGFVEKWNKWTAKETELQEMHEMAQKTAELQEQEL	2735	
QY	64	NTQCGDGVLDLSTKTSIKSSESSICDPSSNS--VAGRLHRNRDDYVYVERSAEFADGL---	119	
DB	2736	SGEKNRAGELQLLEBIKSSQDKQLKELTLENSELKSLDCMDHKDQVEKSGKVEEIAEY	2795	
QY	120	---LSKALKDITOSGALDINKAGIILYGPQTKTLLHLLEALPAGKSPASFNKTRDRPHDSYSY	176	
DB	2796	QLRLHEAEKHKHQAALLDTNKO---YEVEIQT---	2828	
QY	177	KDSKETCAVLQKQVALWARAQARTEKSKLN--LLETSEIKFPTASTYVLHQLTLQKMYTQF	234	
DB	2829	-TSKECLUSSQKLEI---DLLKSKSELNNSLKATQILEELKKTQMDNJ---KYVNLQ	2880	
QY	235	KEKNESL-----QYTESNPVTQLKIPOLRVSSVSKSQPDGSGLLDVMYQVSKT	282	
DB	2881	KKENERAQGMKLLIKSKCKOLEEKEILQKELSQAQAQ-----EKQKT	2924	
QY	283	SSVLE-----GSALQKJKNLTPKONKIE-----CSGPVTHSSVD--SYFLHGDLSPL	327	
DB	2925	GTVMDTKVDLTTBIKELKTELEKTKTEADEYLDKYCSLLISHEKLEKAKEMLETVAAHL	2984	
QY	328	C-----LNSKNGTVDG-----TSENTEGLDRKDSQKPRK-----RGYRQ	364	
DB	2985	CSQSKQDSRSGPLLPVPVPGSPFPSVTEKRLSSGQNKASGKRQRSGIWEANGRGPTPA	3044	
QY	365	YDHEIMBEAIAVMVSGMSVSKAOGI---YGVPH-----STLEYKVKERSGT	408	
DB	3045	TPESFSKSKKAVMSGIHPAEDTGTGTEPEGLVEVVKGFADIPCTGKTSFYILRRITMA	3104	
QY	409	LKTPPK---KKLRUPDTGLVNMVDSGTGCKNNSKP	441	
DB	3105	TRTSPRLAAQKALSLPLSL-----GKENLAESSKP	3134	

  

RESULT 4	
US-09-538-092-184	
Sequence 184, Application US/09538092	
Patent No. 6753314	
GENERAL INFORMATION:	
APPLICANT: Giot, Ioic	
APPLICANT: Mansfield, Traci A.	
TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same	
FILE REFERENCE: 15966-542	
CURRENT APPLICATION NUMBER: US/09/538,092	
CURRENT FILING DATE: 2000-03-29	
PRIOR APPLICATION NUMBER: 60/127,352	
PRIOR FILING DATE: 1999-04-01	
PRIOR APPLICATION NUMBER: 60/179,965	
PRIOR FILING DATE: 2000-02-01	
NUMBER OF SEQ ID NOS: 1387	
SOFTWARE: CuratPatSeqFormatter Version 0.9	
SEQ ID NO 184	
LENGTH: 534	
TYPE: PRT	
ORGANISM: Saccharomyces cerevisiae	
FEATURE:	
NAME/KEY: misc_feature	
LOCATION: (0)...(0)	
OTHER INFORMATION: Polypeptide Accession Number YDR299W	
US-09-538-092-184	

  

Query Match	5.6%;	Score 127;	DB 4;	Length 534;
Best Local Similarity	21.4%;	Pred. No. 0.0052;		
Matches 92;	Conservative 61;	Mismatches 157;	Indels 120;	Gaps 16;

  

QY	5	IRQFAIEYISKSGTKQENRNGSI-----GPSIVCKSIQMNQAENSLQESGQLDLTV	57
----	---	--	----

```

Db      9 ISDIAIPVANKFDIEEENASLFOHNEKNGES-----DLSDYGNSNTEETKKAHYLEV 62
QY     58 NRMQEONTQCGDGLDLSTKK-TGSKSESSICDPSSSENSVAGRLHRNRREDYVSAEP- 115
Db     63 -----EKSKLRAEKGLELNDPKYTGVGSRQALYEEVSENEDEEEEEEEEDALSFR 118
QY    116 -----ADG-----LLSKAL-----KDIQSGALDINKAGI 139
Db    119 TQSEDEVEIDEESDADGGETEBAQQRHAKSLIIQOETKQAINKLSQSVQRDASKG-- 176
QY    140 LYGIPOKTLI-----LHLEALPAGKPAKFNKTRDFHDSYSYKDSKETCAVLQKVALWA 193
Db    177 -YSILOQTUFDNTIDIRILQKAVIAANKLPLTTESWEEAKMDDSEETKRLKK----- 229
QY    194 RQAARTYKSKNLNLETSEIKF-----PTASTYVLHQLTLQKMTVQFKEKNESLOYETS 246
Db    230 --ENEKLFNNLFNRLNFRIKFQLGDHTQNEEVAKHKLKKRSLKELYOETNSLDSELK 287
QY    247 N-PTVQUKIPQLRVSSVSKSQPDGSL-----LDVMYQVSKTSSVLEGSALQKLKN 296
Db    288 EYRTAVLNKWTQVSSASGNAALSSNFKAINLPADVQVENQOLSDMSRLMKRTKLNRRN 346
QY    297 ILPKQNKIECS-----GPVTHSSVDSYFLHGDLSPLCLNSKNGTVDGTSENTEDGLD 348
Db    347 ITFLYFOKOCANGRLPELISPVVKRSDVD-----NENSDDGLD 384
QY    349 RKDSQKPRKK 358
Db    385 IPKNYDPRKK 394

RESULT 5
US-09-248-796A-19154
; Sequence 19154, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 19154
; LENGTH: 1753
; TYPE: PRT
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (223)
; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unknown
US-09-248-796A-19154

Query Match          5.6%; Score 127; DB 4; Length 1753;
Best Local Similarity 21.4%; Pred. No. 0.033;
Matches 95; Conservative 76; Mismatches 156; Indels 116; Gaps 24;

QY    12 YISKSGKGTQENRNGSIGPSIVCKSIQMNQAESLQEOEGPLDLTVNRMQEONTQO---G 68
Db   1342 FITDSGK-ENDD-----EDVEMEDASEVEEEEEEGE-DHEESLSEEEKEEDDDDD 1390
QY    69 DGVLDLSTKTSIKSESSICDPSSSENSVAGRLHRNRREDYV-----ERSAE 114
Db   1391 DDIMDVDDNDENDEDEEQSADESEKSLSTGLSTNGFDWTASILDQADEDESSDDDDTD 1450
QY   115 PADGLLSKALKDIQSGALDINKAGIL-YGIPOKTLILLHLEALPAGKP-----ASPK 164
Db  1451 FTSLKKKSKKKSKSAVED--KTGDNLNTRAPQST--ADFERLLIGNPSSILLMNYISFQ 1506

```



QY 315 VDSYFLHGLDLPCLNSKNGTVDGTSNTDGLDRD-----SKQPRKRGYRQYDHE 368  
Db 859 -----FLRSNLELKEKEDTSTWNYNQKKAASLFEKQLETKSNYKKMEAD 905  
QY 369 IMBEAIAVMVSGKMSVSKAGIYGVPHS-----TLEYKVKERSGTLPKPKKK 416  
Db 906 LQKE-----LQSAFNEINYLNGLAGKVPDRLLSRVELEKVKVSFQKLEKALEEK 956

RESULT 8  
US-08-328-254-6  
; Sequence 6, Application US/08328254  
; Patent No. 5710022  
; GENERAL INFORMATION:  
; APPLICANT: Zhu, Xueliang  
; APPLICANT: Lee, Wen-Hwa  
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/328,254  
; FILING DATE: 24-OCT-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/141,239  
; FILING DATE: 22-OCT-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-CJ 1191  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2482 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-328-254-6

Query Match 5.5%; Score 123; DB 1; Length 2482;  
Best Local Similarity 20.0%; Pred. No. 0.13;  
Matches 103; Conservative 74; Mismatches 201; Indels 138; Gaps 21;

QY 7 QFAIEYISKSGKTOENR---NGSIGPSIVCKSTQMNQAEISLQEEQEGPLDLTVNRMQEQ 63  
Db 1948 QDTLEVLQSSYKNLENELELTQMDKMSFVEKYNKMTAKETELQREHMAKTAELQEEEL 2007  
QY 64 NTQQGDVLDLSTKTSIKSEESSICDPSSENS-VAGRLHRNREDYVERSAEPADGL--- 119  
Db 2008 SGEKNRAGELQALLLEIKSSQDLKELTLENSELKSLDCMHKQDQVEKEGKVEEIAEY 2067  
QY 120 ---LSKALKDIOSGALDINKAGILYGIPOKTLILLHLEALPAGKPAFKNTRDFHDSYSY 176  
Db 2068 QRLHAEAKKHQALLDNTNQ---YEVEIQT-----YREKL----- 2100  
QY 177 KDSKETCAVLQKVALWARAQAERTSKLN--LLETSEIKFPTASTYLTQLTKQMVTOF 234  
Db 2101 -TSKEECLSSQKLEI-----DLKSSKEELNLSKATTQILEELKTKTMDNL---KYVNL 2152  
QY 235 KEKNESL-----QYETSNPTVQLKIPQLRVSSVSKSQPDGSGLLDVMYQVSKT 282

Db 2153 KXENERAQGMKLLIKSKQLEBEKEILOKELSQLQAAQ-----EKQKT 2196  
QY 283 SSYLE-----GSALQKLNILPKQNKIE-----CSGPVTHSSVD--SYFLHGLDLSPL 327  
Db 2197 GTVMDTKVDELTEIKELKETLEETKEADEYLDKYCSLLISHEKLEKAKEMLETQVAHL 2256  
QY 328 C-----LNSKNGTVDG-----TSENTEGGLDRKDSKQPRKK-----RGRYRQ 364  
Db 2257 CSQQSKQDSRSGPILLGPVVPSPVTEKRLSSQGNKASGRQRSSSGIWENGGPPTA 2316  
QY 365 YDHEIMEEAIAVMVSGKMSVSKAAGI-----YGVPH-----STLEYKVKERSGT 408  
Db 2317 TPESFSKSKKXAMVSGIHPAEDTEGTEFEPGEPVVKKGFDADIPGKTSPYLIRRTMA 2376  
QY 409 LKTPPK---KRLRLPDTGLYNNMTDSTGTCCKKSSKP 441  
Db 2377 TRTSPLAAQKLALSPLSL-----GKENLAESSKP 2406

RESULT 9  
US-09-538-092-1154  
; Sequence 1154, Application US/09538092  
; Patent No. 6753314  
; GENERAL INFORMATION:  
; APPLICANT: Giot, Loic  
; APPLICANT: Mansfield, Traci A.  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT APPLICATION NUMBER: US/09/538,092  
; CURRENT FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: 60/127,352  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/178,965  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CurapatSeqformat Version 0.9  
; SEQ ID NO 1154  
; LENGTH: 3210  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Polypeptide Accession Number P49454  
US-09-538-092-1154

Query Match 5.5%; Score 123; DB 4; Length 3210;  
Best Local Similarity 20.0%; Pred. No. 0.2;  
Matches 103; Conservative 74; Mismatches 201; Indels 138; Gaps 21;

QY 7 QFAIEYISKSGKTOENR---NGSIGPSIVCKSTQMNQAEISLQEEQEGPLDLTVNRMQEQ 63  
Db 2676 QDTLEVLQSSYKNLENELELTQMDKMSFVEKYNKMTAKETELQREHMAKTAELQEEEL 2735  
QY 64 NTQQGDVLDLSTKTSIKSEESSICDPSSENS-VAGRLHRNREDYVERSAEPADGL--- 119  
Db 2736 SGEKNRAGELQALLLEIKSSQDLKELTLENSELKSLDCMHKQDQVEKEGKVEEIAEY 2795  
QY 120 ---LSKALKDIOSGALDINKAGILYGIPOKTLILLHLEALPAGKPAFKNTRDFHDSYSY 176  
Db 2796 QRLHAEAKKHQALLDNTNQ---YEVEIQT-----YREKL----- 2828  
QY 177 KDSKETCAVLQKVALWARAQAERTSKLN--LLETSEIKFPTASTYLTQLTKQMVTOF 234  
Db 2829 -TSKEECLSSQKLEI-----DLKSSKEELNLSKATTQILEELKTKTMDNL---KYVNL 2880  
QY 235 KEKNESL-----QYETSNPTVQLKIPQLRVSSVSKSQPDGSGLLDVMYQVSKT 282  
Db 2881 KXENERAQGMKLLIKSKQLEBEKEILOKELSQLQAAQ-----EKQKT 2924  
QY 283 SSYLE-----GSALQKLNILPKQNKIE-----CSGPVTHSSVD--SYFLHGLDLSPL 327

Db 2925 GTVMDTKVDLTTTEIKELKELTLEETKEADEVLDKVCSSLISHEKLEKAKEMLETOVAHL 2984  
QY 328 C-----LNSKNGTVDC-----TSNTEDGLDRKDSKQPRK-----RGRYRQ 364  
Db 2985 CSQSQKDSRSGPGLGPPVPGSPSPVTEKRLSSQNKASGKRQKSSGIWNGGGTPA 3044  
QY 365 YDHEIMEEAIAMVMSGKMSVKAQGI-----YGVPH-----STLEYKVKERSGT 408  
Db 3045 TPESFSKSKKAVMSGIHFAEDTGTGFEPEGLPEVVKGFADIPGTKTSPIYLRRTMA 3104  
QY 409 LKTPPK---KKLRLPTGLYNNMTDSTGSCKNSSKP 441  
Db 3105 TRTSPRLAAQKALSPLSL-----GKNLAESSKP 3134  
  
RESULT 10  
US-09-538-092-399  
; Sequence 399, Application US/09538092  
; Patent No. 6753314  
; GENERAL INFORMATION:  
; APPLICANT: Giot, Loic  
; APPLICANT: Mansfield, Traci A.  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT APPLICATION NUMBER: US/09/538,092  
; CURRENT FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: 60/127,352  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/178,965  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CuratPatSeqformat Version 0.9  
; SEQ ID NO 399  
; LENGTH: 1164  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Polypeptide Accession Number YHR158C  
US-09-538-092-399  
  
Query Match 5.4%; Score 122; DB 4; Length 1164;  
Best Local Similarity 19.7%; Pred. No. 0.05;  
Matches 85; Conservative 69; Mismatches 159; Indels 118; Gaps 17;  
  
QY 7 QFAIEYISKSGKTQENRNGSIGPSIVCKSIQMQAENSLQEQSGPLDLTVNRMQEQTQ 66  
Db 694 QFKIKHYNESSELSON-----NTEIDKLSF-----PVDITIKKSDTAGHD 733  
  
QY 67 QGDGVLDTYK-----TSIKSESSICDPSSSENSVAGRLHNRNREDYVERSABFAD 117  
Db 734 SANHVIDASDEKYNVPMGDVPTDTKNBEASV--PINRDAT-----EVVDEA----- 778  
  
QY 118 GLSKALKALDIQSGALDINKAGILYGIPOKTLHLLEALPAGKPFKNKTRDFHDSYSYK 177  
Db 779 -LFEKLRSELQS-----LXELTHEKALEAG--AHIKELETWOLKLSQK 819  
  
QY 178 DSKEITCAV--LOKVALWARAQARTSKLNLLETSEIKFPTASTYLHOLTQKMTQFKEK 237  
Db 820 NSGTTEIDELDSVRL-----QSKCEILEADNHSLEDKYNVELEFVNSKFLDIEN 869  
  
QY 231 ---VTOFK--EKNESIQVETSPTVOLKIPQLRVSVSKSQPDGGLLDVMYQVSKTSV 286  
Db 870 LNEVIQFONEKIKSLSELE---PNYKELEELQIEHNLSEN----- 908  
  
QY 287 EGSALQKLKILPKONKIECSGPVTHS-----SVDSYFLHGLDPLCLNSKNGTVDGTSN 342  
Db 909 -----ERLKNESKQHNEDIINNANVYSSQLGSLISHWKENRANGSFLSSSSLISVSDEN 963  
  
QY 343 TEDGLDRKDSKQPRKR---GRYQYDHEIMEEAIAMVMSGKMSVKAQGIYGVPHSTLE 399  
Db 964 GEKTVGEPYGDQSRHRRVINKLTNRDLDDLSERSQELTIS--KEKLSSEYHALKMEHSSLS 1022

QY 400 YKVKERSGTLK 410  
Db 1023 QDVLVKEKEIRK 1033  
  
RESULT 11  
US-09-949-016-6417  
; Sequence 6417, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: C1001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6417  
; LENGTH: 1938  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-6417  
  
Query Match 5.4%; Score 122; DB 4; Length 1938;  
Best Local Similarity 17.3%; Pred. No. 0.11;  
Matches 80; Conservative 87; Mismatches 185; Indels 110; Gaps 16;  
  
QY 3 KMIROFAIEYISKSGKTQENRNGSIGPSIVCKSIQMQAENSLQEQSGPLDLTVNRMQE 62  
Db 916 KILLEAKVKELTERLEEEEMNS-----ELVAKRNLEDKCSLKKDID-DLELTITKVEK 970  
  
QY 63 QNTQQQDGVLDLSTKTKTSIKSESSICDPSSSENSVAGRLHNRNREDYVERSABFADGLLSK 122  
Db 971 EKHAATENKYNLSEWTAL---EENISKLTKEKSLQEAHQOTLDDLQVEEDKVNGLIKI 1027  
  
QY 123 ALK-----DIQSGALDINKAGILYGIPOKTLHLLEALPAGKPFKNKTRDFHDSYSYK 177  
Db 1028 NAKLEQOTDDELSLE-----QEKKLADLERAKRLEGLKMSQESIMDLNEK 1077  
  
QY 178 DSKEITCAVLOKVALWARAQARTSKLNLLETSEIKFPTASTYLHOLTQKMTQFKEK 237  
Db 1078 QQIE-----EKLKKEFELSOLQARIDDEQVHSLQFOKKIKELQAR 1118  
  
QY 238 NESLOVE-TSNPTVOLKIPQLR-----VSSVSKSQPDGGLLDVMYQVSKTSV----- 285  
Db 1119 IEELBEEIEAEHTLRAKIEKQSDLALELEIEISERLEEASGATSAQIEMNKKREAFQOM 1178  
  
QY 286 ---LEGASLO-----KLKNILPKONKIECSGPVTHSVDYSYFLH 321  
Db 1179 RRDLEATIQHEATAATLRKKQADSVAELGEQIDNLRVQKLEKEKSELKMEIDD--NA 1236  
  
QY 322 GDLSPLCLNSKNGTVDGTSNTED-----GLDRKDSKQPRKRGRYQYDHEI-MEEAIAMV 377  
Db 1237 SNTLEALSCKSN--IERTCTVEDQSEIKANDEQ-----TOLIHLDNNQKARLOT 1286  
  
QY 378 MSGKMSVKAQGIYGVPHSTLEYKVKERSGTLPKPKKLRL 419  
Db 1287 QNGELS-----HRVEEKESLISQLTKSKOAL 1312  
  
RESULT 12  
US-09-949-016-8134  
; Sequence 8134, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:





101 LHRNREDDYVERSAEADGGLSKALKDIOGSGALDINKAGILYIPOKTL--LHLEALPAG 158  
1032 FHLRTSYKKTSSKLDGLLERRIKQ-----FTLEBKORLEKIKLEGGIKG 1077  
159 KPASFKNKTRDFHDSYSYKDSKETC---AVLQKVALWAPAQAERTE-----KSKLNL 208  
1078 IGTSTNSKNLSSESPVITKALKEGOSDSMQEQS--PNANNQPEDLLOQCSQSDSVL 1135  
209 ETSEIKFPASTY-----LHQLTQKMWTFQKEKNELOQYETSNPTVOLKIPOL--RVSS 261  
1136 RMSDPSTHTNKLYPKDRVLDDVSIRSPTKPKQN-SIEND-----IEKVSDLASRGQE 1189  
262 VSKSQPDGGLLDVMYQVSKTSVLE-GSALQKLNILPKQNK---IECSGPVTHSSVDS 317  
1190 PTKSKTKGN---DFFIDDSKLASADDITGLCKNKKPLIQESDITIVSSSKSALHSSVPK 1246  
318 YFLHGLDSP-----CLNSKNGTVDGTS-----ENTDGLDRKDKQPKRKG 360  
1247 STNDRDATPLSRAMDFEGKLGCDSESNSTLSESDTVSIQDSSEDMIVQNSNESISEQF 1306  
361 RYROYDHEIMEEAIAMVMSGKMSVSKAQGIYGVPHSTLEYKVKERSGTLKTPPKKKLRIP 420  
1307 RTHQDVEVLEPLKCELVSGEST-----GNCEDRLPVKGTANGKKPSQOKKLEERP 1358  
421 -----DTGLYNMTDSTGSGCKNSSK 440  
1359 VNKCSQIKLKNNTTDDKNNENRESEK 1384

RESULT 15  
US-09-248-796A-14568  
; Sequence 14568, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 14568  
; LENGTH: 475  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-14568

Query Match 5.3%; Score 119.5; DB 4; Length 475;  
Best Local Similarity 22.3%; Pred. No. 0.021;  
Matches 96; Conservative 55; Mismatches 180; Indels 99; Gaps 16;

39 NQAEENSLQEQSGPLDLTVN--RMQEQNTQOGDGLDLSTKTKTSIKSESSICDPSPSENS 96  
88 SQTQPSLQKRDSPVPGAVSSPNPTPSQPSGERSLSQRAPS-SSNDDLADPA--- 142  
97 VAGRLHRNREDYVERSAEPADGGLSKALKDIOSGALDINKAGILYIPOKTLHLLEALP 156  
143 -----ISGOLSQATSDI-----ANVSNQIKSLTSQTNLHEKTR 177  
157 AGK-----PASFNKTRDFHDSYKDSKETCAVLQKVALWARAQAE-----RTEKSK 204  
178 AEKELQRLITKSEIENKILQRLASY-----DNEVIQEQVEANLATAKEETEALRSEAS- 232  
205 LNLLETSEIKFPTASTYHLQL-----TLQKMWTFQKEKNELOQYETSNPTVOLKIPQLRV 259  
233 -----IAEAKVNSLSGELHEKQVAMESLOKENSTLKEKLSNAE-----IVELEK 278  
260 SVSVKSPQDGSGLLDVMYQVSKTSVLEGSALQKLNILPKQNKIECSGPVTHSSVDSYF 319

279 QAASKSOETHA-----LSNQVAVKKSQVQ-VAVIVKSELKSKIAEIEASHKQLQLDLD--- 330  
320 LHGDLSPCLNSKNGTVDGTSENTEDGLDRK-----DSKOPRKRGRYQYDHEIMEEAI 375  
331 -----NAERERLD--SENRAKELHAKSVLELQNKPTKPSAGITATIGAAAVGAAA 378  
376 MWMSGKMSV-----SKAQGIYGVPHSTLEYKVKERSGTLKTPPKKKLRPLPDTGLYNMT 428  
379 VVGGVASVVHLSAGSTESSDSGRNDPVERSVKETEESDNDVEETNARFPQLQVSEPA 438  
429 DSGTGCKNS 438  
439 DNATNTTSSN 448

Search completed: February 24, 2005, 01:55:57  
Job time : 46 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 24, 2005, 01:54:30 ; Search time 131 Seconds  
(without alignments)  
1104.126 Million cell updates/sec

Title: US-10-016-768A-8  
Perfect score: 2250  
Sequence: 1 MKKMTQFAIEYISKGTQ.....GLYNTDSTGCKNSKRPV 442

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1380268 seqs, 327241040 residues

Total number of hits satisfying chosen parameters: 1380268

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA.\*

- 1: /cgn2\_6/prodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/prodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/prodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/prodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/prodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/prodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/prodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/prodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/prodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/prodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/prodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/prodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/prodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/prodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/prodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/prodata/1/pubpaa/US10D\_PUBCOMB.pep.\*
- 17: /cgn2\_6/prodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/prodata/1/pubpaa/US11\_NEW\_PUB.pep.\*
- 19: /cgn2\_6/prodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 20: /cgn2\_6/prodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2250	100.0	442	13	US-10-016-768-8
2	1754	78.0	357	14	US-10-085-198-24
3	1411	62.7	277	15	US-10-029-386-33895
4	1090	48.4	213	15	US-10-094-749-1949
5	502	22.3	311	15	US-10-108-260A-3561
6	499.5	22.2	572	16	US-10-408-785A-1766
7	294.5	13.1	104	11	US-09-864-408A-2848
8	273	12.1	53	13	US-10-016-768-2
9	233.5	10.4	54	13	US-10-016-768-3
10	229	10.2	53	13	US-10-016-768-4
11	200.5	8.9	1165	13	US-10-016-768-10
12	165	7.3	53	13	US-10-016-768-1
13	163	7.2	53	13	US-10-016-768-5

14	133.5	5.9	1109	16	US-10-387-767-8	Sequence 8, Appli
15	132.5	5.9	848	13	US-10-011-588-45	Sequence 45, Appli
16	132	5.9	787	16	US-10-437-963-125796	Sequence 125796,
17	131.5	5.8	870	14	US-10-029-386-32827	Sequence 32827, A
18	131.5	5.8	2375	16	US-10-408-765A-277	Sequence 277, App
19	131.5	5.8	2492	16	US-10-697-526-2	Sequence 2, Appli
20	131	5.8	1297	15	US-10-354-774-77	Sequence 77, Appl
21	131	5.8	1297	15	US-10-271-012-77	Sequence 77, Appl
22	131	5.8	1297	15	US-10-452-024-7	Sequence 7, Appli
23	131	5.8	1297	15	US-10-452-024-150	Sequence 150, App
24	131	5.8	1297	15	US-10-205-516-14	Sequence 14, Appl
25	131	5.8	1297	16	US-10-729-122-77	Sequence 77, Appl
26	131	5.8	1297	16	US-10-729-039-77	Sequence 77, Appl
27	131	5.8	2017	15	US-10-452-024-146	Sequence 146, App
28	127	5.6	534	10	US-09-893-519A-37	Sequence 37, Appl
29	126	5.6	433	15	US-10-369-493-2477	Sequence 2477, Ap
30	125.5	5.6	1819	15	US-10-335-977-7981	Sequence 7981, Ap
31	125.5	5.6	1820	15	US-10-335-977-7982	Sequence 7982, Ap
32	125	5.6	1307	15	US-10-205-516-28	Sequence 28, Appl
33	124.5	5.5	972	9	US-09-924-154-16	Sequence 16, Appl
34	123.5	5.5	1827	15	US-10-369-493-5368	Sequence 5368, Ap
35	123	5.5	1017	16	US-10-408-765A-158	Sequence 158, App
36	122.5	5.4	748	15	US-10-425-114-72422	Sequence 72422, A
37	122.5	5.4	958	15	US-10-369-493-6600	Sequence 6600, Ap
38	122.5	5.4	958	15	US-10-369-493-6601	Sequence 6601, Ap
39	122	5.4	454	15	US-10-369-339-70	Sequence 70, Appl
40	122	5.4	454	15	US-10-369-100-160	Sequence 160, App
41	122	5.4	1435	15	US-10-369-493-1390	Sequence 1390, Ap
42	122	5.4	1938	16	US-10-408-765A-1168	Sequence 1168, Ap
43	122	5.4	1938	16	US-10-408-765A-1629	Sequence 1629, Ap
44	122	5.4	2781	15	US-10-263-929-122	Sequence 122, App
45	121.5	5.4	1384	16	US-10-473-576-22	Sequence 22, Appl

ALIGNMENTS

RESULT 1  
US-10-016-768-8  
; Sequence 8, Application US/10016768  
; Publication No. US20020142443A1  
; GENERAL INFORMATION:  
; APPLICANT: Baehrcke, Eric H.  
; TITLE OF INVENTION: GENES REGULATING PROGRAMMED CELL DEATH  
; FILE REFERENCE: 4115-131  
; CURRENT APPLICATION NUMBER: US/10/016,768  
; CURRENT FILING DATE: 2001-10-29  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 8  
; LENGTH: 442  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-016-768-8

Query Match 100.0%; Score 2250; DB 13; Length 442;  
Best Local Similarity 100.0%; Pred. No. 2e-169;  
Matches 442; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MKKMTQFAIEYISKGTQENRNGSIGPSIVCKSTQMNQAEISLOEQEGPLDLTVNRM	60
DB	1	MKKMTQFAIEYISKGTQENRNGSIGPSIVCKSTQMNQAEISLOEQEGPLDLTVNRM	60
QY	61	QEQNTQQQGVLDLSTKTSIKSESSICDPSENSVAGLRHNRREDYVERSAEFADGLL	120
DB	61	QEQNTQQQGVLDLSTKTSIKSESSICDPSENSVAGLRHNRREDYVERSAEFADGLL	120
QY	121	SKALXDIQSGALDINKAGILYGIPOKTLHLLEALPAGKPFKNTKTRDFHDSYSYKDSK	180
DB	121	SKALXDIQSGALDINKAGILYGIPOKTLHLLEALPAGKPFKNTKTRDFHDSYSYKDSK	180
QY	181	ETCAVLQKVALWARQAERTKSKNLLTSEIKPTASTYLHOLTLOKMTQFKENES	240

Db 181 ETCVALQKVALWARAQARTKSKLNLETSEIKPTASTYHLQTLQKMTQFKKNES 240  
Qy 241 LOYETSNPTVLKIPOLRVSSVSKSQPDGSLDWMYQVSTSSVLEGSALQKLNILPK 300  
Db 241 LOYETSNPTVLKIPOLRVSSVSKSQPDGSLDWMYQVSTSSVLEGSALQKLNILPK 300  
Qy 301 QNKIECSGPTVTHSSVDSYFLHGDLSPLCLNSKNGTVDTGSENTEGLDRKDSKQPRKRG 360  
Db 301 QNKIECSGPTVTHSSVDSYFLHGDLSPLCLNSKNGTVDTGSENTEGLDRKDSKQPRKRG 360  
Qy 361 RYQYDHEIMEBAIAMVMSGKMSVSKAQGIYGVPHSTLEYKVKERSGTLTKTPPKKRLP 420  
Db 361 RYQYDHEIMEBAIAMVMSGKMSVSKAQGIYGVPHSTLEYKVKERSGTLTKTPPKKRLP 420  
Qy 421 DTGLYNMTDSGTGCKNSKPV 442  
Db 421 DTGLYNMTDSGTGCKNSKPV 442

RESULT 2

US-10-085-198-24  
; Sequence 24, Application US/10085198  
; Publication No. US20040009907A1  
; GENERAL INFORMATION:  
; APPLICANT: Alsobrook et al.  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-279  
; CURRENT APPLICATION NUMBER: US/10/085,198  
; CURRENT FILING DATE: 2002-02-25  
; PRIOR APPLICATION NUMBER: 60/271,646  
; PRIOR FILING DATE: 2001-02-26  
; PRIOR APPLICATION NUMBER: 60/276,401  
; PRIOR FILING DATE: 2001-03-16  
; PRIOR APPLICATION NUMBER: 60/311,981  
; PRIOR FILING DATE: 2001-08-13  
; PRIOR APPLICATION NUMBER: 60/312,858  
; PRIOR FILING DATE: 2001-08-16  
; PRIOR APPLICATION NUMBER: 60/271,840  
; PRIOR FILING DATE: 2001-02-27  
; PRIOR APPLICATION NUMBER: 60/277,324  
; PRIOR FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: 60/286,096  
; PRIOR FILING DATE: 2001-04-21  
; PRIOR APPLICATION NUMBER: 60/299,695  
; PRIOR FILING DATE: 2001-06-20  
; PRIOR APPLICATION NUMBER: 60/315,614  
; PRIOR FILING DATE: 2001-08-29  
; PRIOR APPLICATION NUMBER: 60/272,405  
; PRIOR FILING DATE: 2001-02-28  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 653  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 24  
; LENGTH: 356  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-085-198-24

Query Match 78.0%; Score 1754; DB 15; Length 356;  
Best Local Similarity 100.0%; Pred. No. 2.6e-130; Indels 0; Gaps 0;  
Matches 343; Conservative 0; Mismatches 0

Qy 100 RLHRNREYVERSAEFADGLLSKALKDIQSGALDINKAGILYGIPOKTLHLHLEALPAGK 159  
Db 14 RLHRNREYVERSAEFADGLLSKALKDIQSGALDINKAGILYGIPOKTLHLHLEALPAGK 73

Qy 160 PASFNKTRDFHDSYSYKDSKETCAVLQKVALWARAQARTKSKLNLETSEIKFPTAS 219  
Db 74 PASFNKTRDFHDSYSYKDSKETCAVLQKVALWARAQARTKSKLNLETSEIKFPTAS 133

Qy 220 TYLHQLTLQKMTQFKKNESLOYESNPTVOLKIPOLRVSSVSKSQPDGSLDWMYQV 279  
Db 134 TYLHQLTLQKMTQFKKNESLOYESNPTVOLKIPOLRVSSVSKSQPDGSLDWMYQV 193

Qy 280 SKTSSVLEGSALQKLNILPKQNKIECSGPTVTHSSVDSYFLHGDLSPLCLNSKNGTVDTG 339  
Db 194 SKTSSVLEGSALQKLNILPKQNKIECSGPTVTHSSVDSYFLHGDLSPLCLNSKNGTVDTG 253  
Qy 340 SENTEDGLDRKDSKQPRKRGRYQYDHEIMEBAIAMVMSGKMSVSKAQGIYGVPHSTLE 399  
Db 254 SENTEDGLDRKDSKQPRKRGRYQYDHEIMEBAIAMVMSGKMSVSKAQGIYGVPHSTLE 313  
Qy 400 YKVKERSGTLTKTPPKKRLPDTGLYNMTDSGTGCKNSKPV 442  
Db 314 YKVKERSGTLTKTPPKKRLPDTGLYNMTDSGTGCKNSKPV 356

RESULT 3

US-10-029-386-33895  
; Sequence 33895, Application US/10029386  
; Publication No. US20030194704A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GI  
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO  
; FILE REFERENCE: AEONICA-X-2  
; CURRENT APPLICATION NUMBER: US/10/029,386  
; CURRENT FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 34288  
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 33895  
; LENGTH: 277  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC005768.16  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.85  
; OTHER INFORMATION: SWISSPROT HIT: Q9YID8, EVALUATE 1.60e+00  
US-10-029-386-33895

Query Match 62.7%; Score 1411; DB 14; Length 277;  
Best Local Similarity 100.0%; Pred. No. 2.5e-103; Indels 0; Gaps 0;  
Matches 277; Conservative 0; Mismatches 0

Qy 100 RLHRNREYVERSAEFADGLLSKALKDIQSGALDINKAGILYGIPOKTLHLHLEALPAGK 159  
Db 1 RLHRNREYVERSAEFADGLLSKALKDIQSGALDINKAGILYGIPOKTLHLHLEALPAGK 60

Qy 160 PASFNKTRDFHDSYSYKDSKETCAVLQKVALWARAQARTKSKLNLETSEIKFPTAS 219  
Db 61 PASFNKTRDFHDSYSYKDSKETCAVLQKVALWARAQARTKSKLNLETSEIKFPTAS 120

Qy 220 TYLHQLTLQKMTQFKKNESLOYESNPTVOLKIPOLRVSSVSKSQPDGSLDWMYQV 279  
Db 121 TYLHQLTLQKMTQFKKNESLOYESNPTVOLKIPOLRVSSVSKSQPDGSLDWMYQV 180

Qy 280 SKTSSVLEGSALQKLNILPKQNKIECSGPTVTHSSVDSYFLHGDLSPLCLNSKNGTVDTG 339  
Db 181 SKTSSVLEGSALQKLNILPKQNKIECSGPTVTHSSVDSYFLHGDLSPLCLNSKNGTVDTG 240

Qy 340 SENTEDGLDRKDSKQPRKRGRYQYDHEIMEBAIAM 376  
Db 241 SENTEDGLDRKDSKQPRKRGRYQYDHEIMEBAIAM 277

RESULT 4

US-10-094-749-1949  
; Sequence 1949, Application US/10094749  
; Publication No. US20030219741A1  
; GENERAL INFORMATION:  
; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: SUGIYAMA, TOMOYASU  
; APPLICANT: OTSUKI, TETSUJI  
; APPLICANT: WAKAMATSU, AI

APPLICANT: SATO, HIROYUKI  
APPLICANT: ISHII, SHIZUKO  
APPLICANT: YAMAMOTO, JUN-ICHI  
APPLICANT: ISONO, YUUKO  
APPLICANT: HIO, YURI  
APPLICANT: OTSUKA, KAORU  
APPLICANT: NAGAI, KEIICHI  
APPLICANT: TRIE, RYOTARO  
APPLICANT: TAMECHIKI, ICHIRO  
APPLICANT: SEKI, NAOHIKO  
APPLICANT: YOSHIKAWA, TSUTOMU  
APPLICANT: OTSUKA, MOTYUKI  
APPLICANT: NAGAHARI, KENJI  
APPLICANT: MASUHO, YASUHIKO  
TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA  
FILE REFERENCE: 084335/0160  
CURRENT APPLICATION NUMBER: US/10/094,749  
CURRENT FILING DATE: 2002-03-12  
PRIOR APPLICATION NUMBER: 60/350,435  
PRIOR FILING DATE: 2002-01-24  
PRIOR APPLICATION NUMBER: JP 2001-328381  
PRIOR FILING DATE: 2001-09-14  
NUMBER OF SEQ ID NOS: 3381  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1949  
LENGTH: 213  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-094-749-1949

Query Match 48.4%; Score 1090; DB 15; Length 213;  
Best Local Similarity 99.5%; Pred. No. 4.2e-78;  
Matches 212; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 230 MTQPKKNESLQYETSNPTVOLKIPQLRVSSVSKSQPDGSLDVMYQVSKTSSVLEGS 289  
DB 1 MTQPKKNESLQYETSSFTVQKIPQLRVSSVSKSQPDGSLDVMYQVSKTSSVLEGS 60  
QY 290 ALQKLNILPKQNKIECSGPVTHSSVDSYFLHGDLSPLCLNSKNGTVDTGSENTEGLDR 349  
DB 61 ALQKLNILPKQNKIECSGPVTHSSVDSYFLHGDLSPLCLNSKNGTVDTGSENTEGLDR 120  
QY 350 KDSQPKRKRGRYQYDHEIMEAIAVMGSKMSVSKAQGIYGVPHSTLEYKVKERSGTL 409  
DB 121 KDSQPKRKRGRYQYDHEIMEAIAVMGSKMSVSKAQGIYGVPHSTLEYKVKERSGTL 180  
QY 410 KTPPKKRLPDTGLNMTDSGTGCKNSKPV 442  
DB 181 KTPPKKRLPDTGLNMTDSGTGCKNSKPV 213

RESULT 5  
US-10-108-260A-3561  
Sequence 3561, Application US/10108260A  
Publication No. US20040005560A1  
GENERAL INFORMATION:  
APPLICANT: HELIX RESEARCH INSTITUTE  
TITLE OF INVENTION: No. US20040005560A1e1 full length cDNA  
FILE REFERENCE: H1-A0106  
CURRENT APPLICATION NUMBER: US/10/108,260A  
CURRENT FILING DATE: 2002-03-27  
NUMBER OF SEQ ID NOS: 5458  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 3561  
LENGTH: 311  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-108-260A-3561

Query Match 22.3%; Score 502; DB 15; Length 311;  
Best Local Similarity 89.5%; Pred. No. 2.4e-31;  
Matches 102; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 MKKMIRQFAIEYISKSGKTOENRNGSIGPSIVCKSIOMNOAENSLOEEOGGLDITVNRM 60  
DB 154 MKKMIRQFAIEYISKSGKTOENRNGSIGPSIVCKSIOMNOAENSLOEEOGGLDITVNRM 213  
QY 61 QEONTQOGDGVLDLSTKTKTSIKSEESSICDPSSSENSVAGRLHNRNREDYVERSAE 114  
DB 214 QEONTQOGDGVLDLSTKTKTSIKSEESSICDPSSSENSVAGRLHNRNREDYVERSAE 267  
RESULT 6  
US-10-408-765A-1766  
Sequence 1766, Application US/10408765A  
Publication No. US20040101874A1  
GENERAL INFORMATION:  
APPLICANT: Ghosh, Soumitra S.  
APPLICANT: Fahy, Eoin D.  
APPLICANT: Zhang, Bing  
APPLICANT: Gibson, Bradford W.  
APPLICANT: Taylor, Steven W.  
APPLICANT: Glenn, Gary M.  
APPLICANT: Warnock, Dale B.  
TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION  
FILE REFERENCE: 660088.465  
CURRENT APPLICATION NUMBER: US/10/408,765A  
CURRENT FILING DATE: 2003-04-04  
NUMBER OF SEQ ID NOS: 3077  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1766  
LENGTH: 572  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-408-765A-1766

Query Match 22.2%; Score 499.5; DB 16; Length 572;  
Best Local Similarity 33.6%; Pred. No. 9.3e-31;  
Matches 158; Conservative 66; Mismatches 143; Indels 103; Gaps 20;

QY 1 MKKMIRQFAIEYISKSGKTOE-----NRNGS-----IGPSIVCKSIOMNOAENSLOEE 48  
DB 140 MQRMICQFAAEYTSKNSSTQDPNSKQSLPKASPVTTSTPTAATTQNPVLSKLLMAD 199  
QY 49 QSGPLDITVNRQEQNTQOGDGVLDLSTKTKT-----SIKSEESSICDPSSSENSVAGRLHNRN 104  
DB 200 QDSPLDITVNRQEQNTQOGDGVLDLSTKTKSPCAGSTLSHSPGCSSTQGNRGRPSQY 258  
QY 105 REDYVERSAEFAFGDGLSKALKD-----IQSGALDINKAGILYGIQKTLILLHLEALPAG 158  
DB 259 RPD-----GLRSGDGVPPRSLQDGTREGFGHSTSLKVPPLA-----RSLOISEELLSRN 306  
QY 159 K-----PASFKNKTRDFHDSYSYKDSKETCAVLQKVALWARAQAE--RTEKSKLN--- 206  
DB 307 QLSTAASLGLSPGLQN-----HQGH-----LILSREASWAKPHVEFNLSRMKFRNG 352  
QY 207 -LLETSEIKFPTASTYILHOLTQKMTQPKKNESLQYETSNPTVOLKIPOLRVSSVSKS 265  
DB 353 ALSNISDLPFLAENSAPFPMALQ-----AKODGKKDYSHSPVDLKIPOVRGMDLSWE 405  
QY 266 QPDGSGLLDVMYQVSKTSSVL-----EGSALQKLNILPKQNKIEC--SGPVTHSSVDSYF 319  
DB 406 SRTGD-----QYSYSLVMGSGQTESALSKLRLAILPKQSRKSMLDADG-----DSWG 452  
QY 320 LHGDLSPLCLNSKNGTVDTGSENTEGLDRKDSQPKRKRGRYQYDHEIMEAIAVMMS 379  
DB 453 SDAE-----QSTSGQPYPTSDQEGD-----PGSKQPKRKRGRYQYNSILEEAIIVMS 502  
QY 380 GKMSVSKAQGIYGVPHSTLEYKVKERSGTLKTPPKKRL-----PDTGL 424  
DB 503 GKMSVSKAQSIYGIPIHSTLEYKVKERLGLTKNPPKKMKLMRSEGPDSV 552  
RESULT 7  
US-09-864-408A-2848

```
; Sequence 2848, Application US/09864408A
; Publication No. US20040009474A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shinketsu, Richard A.
; TITLE OF INVENTION: No. US20040009474A1 Human Polynucleotides and Polypeptides Encod
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2848
; LENGTH: 104
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (24)..(24)
; OTHER INFORMATION: Wherein Xaa may be any naturally occurring amino acid
US-09-864-408A-2848

Query Match      13.1%; Score 294.5; DB 11; Length 104;
Best Local Similarity 74.4%; Pred. No. 1.3e-15;
Matches 58; Conservative 7; Mismatches 8; Indels 5; Gaps 1;

QY 352 SKQPKRGRYQYDHEIMEEAIAMVMSGKMSVSKAQGIYGVPHSTLEYKVGRTLT 411
Db 7 SKQPKRGRYQYDHEIMEEAIAMVMSGKMSVSKAQGIYGVPHSTLEYKVGRTLT 66

QY 412 PPKKLRL-----PDGTL 424
Db 67 PPKKKMLRSEGPVDSV 84

RESULT 8
US-10-016-768-2
; Sequence 2, Application US/10016768
; Publication No. US20020142443A1
; GENERAL INFORMATION:
; APPLICANT: Baehrecke, Eric H.
; TITLE OF INVENTION: GENES REGULATING PROGRAMMED CELL DEATH
; FILE REFERENCE: 4115-131
; CURRENT APPLICATION NUMBER: US/10/016,768
; CURRENT FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(54)
; OTHER INFORMATION: X CAN BE ANY AMINO ACID
US-10-016-768-2

Query Match      12.1%; Score 273; DB 13; Length 53;
Best Local Similarity 100.0%; Pred. No. 2.5e-14;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 353 KQPKRGRYQYDHEIMEEAIAMVMSGKMSVSKAQGIYGVPHSTLEYKVKER 405
Db 1 KQPKRGRYQYDHEIMEEAIAMVMSGKMSVSKAQGIYGVPHSTLEYKVKER 53

RESULT 9
US-10-016-768-3
; Sequence 3, Application US/10016768
; Publication No. US20020142443A1
; GENERAL INFORMATION:
; APPLICANT: Baehrecke, Eric H.
; TITLE OF INVENTION: GENES REGULATING PROGRAMMED CELL DEATH
; FILE REFERENCE: 4115-131
; CURRENT APPLICATION NUMBER: US/10/016,768
; CURRENT FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 1165
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-016-768-3

Query Match      8.9%; Score 200.5; DB 13; Length 1165;
Best Local Similarity 21.7%; Pred. No. 1.2e-06;
Matches 92; Conservative 66; Mismatches 146; Indels 119; Gaps 14;

; TITLE OF INVENTION: GENES REGULATING PROGRAMMED CELL DEATH
; FILE REFERENCE: 4115-131
; CURRENT APPLICATION NUMBER: US/10/016,768
; CURRENT FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 54
; TYPE: PRT
; ORGANISM: T. nigroviridis
US-10-016-768-3

Query Match      10.4%; Score 233.5; DB 13; Length 54;
Best Local Similarity 81.5%; Pred. No. 3.4e-11;
Matches 44; Conservative 7; Mismatches 2; Indels 1; Gaps 1;

QY 353 KQPKRGRYQYDHEIMEEAIAMVMSGKMSVSKAQGIYGVPHSTLEYKVKER 405
Db 1 KQPKRGRYQYDHEIMEEAIAMVMSGKMSVSKAQGIYGVPHSTLEYKVKER 54

RESULT 10
US-10-016-768-4
; Sequence 4, Application US/10016768
; Publication No. US20020142443A1
; GENERAL INFORMATION:
; APPLICANT: Baehrecke, Eric H.
; TITLE OF INVENTION: GENES REGULATING PROGRAMMED CELL DEATH
; FILE REFERENCE: 4115-131
; CURRENT APPLICATION NUMBER: US/10/016,768
; CURRENT FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 53
; TYPE: PRT
; ORGANISM: M. musculus
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(54)
; OTHER INFORMATION: X can be any amino acid
US-10-016-768-4

Query Match      10.2%; Score 229; DB 13; Length 53;
Best Local Similarity 81.1%; Pred. No. 7.5e-11;
Matches 43; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 353 KQPKRGRYQYDHEIMEEAIAMVMSGKMSVSKAQGIYGVPHSTLEYKVKER 405
Db 1 KQPKRGRYQYDHEIMEEAIAMVMSGKMSVSKAQGIYGVPHSTLEYKVKER 53

RESULT 11
US-10-016-768-10
; Sequence 10, Application US/10016768
; Publication No. US20020142443A1
; GENERAL INFORMATION:
; APPLICANT: Baehrecke, Eric H.
; TITLE OF INVENTION: GENES REGULATING PROGRAMMED CELL DEATH
; FILE REFERENCE: 4115-131
; CURRENT APPLICATION NUMBER: US/10/016,768
; CURRENT FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 1165
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-016-768-10

Query Match      8.9%; Score 200.5; DB 13; Length 1165;
Best Local Similarity 21.7%; Pred. No. 1.2e-06;
Matches 92; Conservative 66; Mismatches 146; Indels 119; Gaps 14;
```

```
QY 19 TQNRNGSTGPSTIVCKSIQNMQAENSLQEBQGLDLTVNRMQEQNTQQDGVLDL--ST 76
:||||: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 514 SQENSGNASLLQQOQHQQHQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQO
QY 77 KKTSIKSESSICDPSSSENSVAGRLHRNREDYVERSAEPADGLLSKALDKIQSGALDINK 136
: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 574 EPILKIPKSVGPPASSS-----LSP 596
QY 137 AGILYIPQKTLHLHLALPAGKPFKNKTRDFHDSYSYKO-----SKETCAVLQKVALW 192
: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 597 GGLVGG-----HHPLNNNSLSISNNSN--HSSNSHRNGSNKSPHSASPLAAAV- 645
QY 193 ARAQAERTKSKNLLTSEIKPPTASTYLHQLTQKMTQFKERKNEISQYETSNTVOL 252
: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 646 --AQGYSAAGNSLLTSSSSIQNMASNIQRQI-----NEQSGQES----- 684
QY 253 KIPQLRVSSVSKSQDPGSLDVMYO-----VSKTSSVLEGSALQKLNLPKONKIECS 307
: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 685 ----LRNGVDCSSNNGSSSSLYGKKPSISVAKIIGGTDTSRFGASPNLLSQOH----- 735
QY 308 GPVTHSSVDSYFLHGDLSPLCLNSKNGTVDGTSENTEDGLDRKDS--KQPRKKRGYROY 365
: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 736 ----HS---AHL-----THQOQQOQLSAQALGKGRPKRGYRNY 770
QY 366 DHEIMEEATAMVSGKMSVSKAQGIYGVPHSTLEYKVKERSGTLKTPPKKLRLLPTGLY 425
: : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 771 DRDSLVEAVKAVQRGMSVHRAGSYGVPHSTLEYKVKERH---LMRPKRREPQKQPDILV 827
QY 426 NMT 428
: :
Db 828 GLT 830
```

```
RESULT 12
US-10-016-768-1
; Sequence 1, Application US/10016768
; Publication No. US20020142443A1
; GENERAL INFORMATION:
; APPLICANT: Baehrecke, Eric H.
; TITLE OF INVENTION: GENES REGULATING PROGRAMMED CELL DEATH
; FILE REFERENCE: 4115-131
; CURRENT APPLICATION NUMBER: US/10/016,768
; CURRENT FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (1)..(54)
; OTHER INFORMATION: X can be any amino acid
US-10-016-768-1
```

```
Query Match 7.3%; Score 165; DB 13; Length 53;
Best Local Similarity 60.4%; Pred. No. 8.7e-06;
Matches 32; Conservative 6; Mismatches 15; Indels 0; Gaps 0;
```

```
QY 353 KQPRKKRGYROYDHEIMEEATAMVSGKMSVSKAQGIYGVPHSTLEYKVKER 405
Db 1 KGTTPRKGRYRNYDRDSLVEAVKAVQRGMSVHRAGSYGVPHSTLEYKVKER 53
```

```
RESULT 13
US-10-016-768-5
; Sequence 5, Application US/10016768
; Publication No. US20020142443A1
; GENERAL INFORMATION:
; APPLICANT: Baehrecke, Eric H.
; TITLE OF INVENTION: GENES REGULATING PROGRAMMED CELL DEATH
; FILE REFERENCE: 4115-131
```

```
; CURRENT APPLICATION NUMBER: US/10/016,768
; CURRENT FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (1)..(54)
; OTHER INFORMATION: X CAN BE ANY AMINO ACID
US-10-016-768-5
```

```
Query Match 7.2%; Score 163; DB 13; Length 53;
Best Local Similarity 56.6%; Pred. No. 1.2e-05;
Matches 30; Conservative 10; Mismatches 13; Indels 0; Gaps 0;
```

```
QY 353 KQPRKKRGYROYDHEIMEEATAMVSGKMSVSKAQGIYGVPHSTLEYKVKER 405
Db 1 KRSRPRGQYRKYDKNALDEAVRSVRRGEMTVHRAGSFFGVPHSTLEYKVKER 53
```

```
RESULT 14
US-10-387-767-8
; Sequence 8, Application US/10387767
; Publication No. US20040180845A1
; GENERAL INFORMATION:
; APPLICANT: Newgard, Christopher B.
; APPLICANT: Yang, Ruojing
; TITLE OF INVENTION: GLYCEN TARGETING SUBUNITS OF PROTEIN PHOSPHATASE
; FILE REFERENCE: 5405-302
; CURRENT APPLICATION NUMBER: US/10/387,767
; CURRENT FILING DATE: 2003-03-13
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8
; LENGTH: 1109
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-10-387-767-8
```

```
Query Match 5.9%; Score 133.5; DB 16; Length 1109;
Best Local Similarity 22.9%; Pred. No. 0.22;
Matches 89; Conservative 53; Mismatches 163; Indels 83; Gaps 17;
```

```
QY 7 QFAIEYISKSGKTQENRNGSIGPSIVCKSIQNMQAENSLQEBQGLDLTVNRMQEQNTQ 66
Db 208 EFCIRYETSVGTFWNNNGT-NYTLVCQ-----KKEPEPEPKPL-----EAPSK 232
QY 67 QGDGVLDLSTKRTSIKSESSICDPSSSENSVAGRLHRNREDYVER---SAEFADGLLS-- 121
Db 253 QKKGCLKVKSSK-----BESS--ETSEENNFE--SKIADTYIPTIVCSHEEKEDLKSSY 303
QY 122 KALKDIQSGALDINKAGILYIPQKTLHLHLALPAGK-----PASFKNKTDFHDSYS 175
Db 304 QNVKDVNTEHDEHNEKELEMINORLIRTRCAASEYKQNTLSSDPSNIPKPELOKNQS 363
QY 176 YKDSKETCAVLQKVALWARAQARTKSKNLLTSEIKPPTASTYLHQLTQKMTQPK 235
Db 364 H---SEACTDLSQLRLSPGSSAESLKGDFYHTE-----KYSGNESSHQ-----PSDMG 410
QY 236 EKNEISQYETSNTVOLKIPQLRVSSVSKSQDPGSLDVL-----MYQVSKTSSVLEGS 289
Db 411 EINPSLGGTTSDGSVQLHISSKKEILLDDNANPAHSGRGEISCSFPQQLKASLNKKYEGG 470
QY 290 ALQKLNILPKQNKIECSGPVTHSSVDSYFLHGDLSPLCLNSKNGTVDGTSENTEDGLDR 349
Db 471 AENS-----EMKDCBCLPRDVHLKASDYF-----KKSTENRPSE--EDYGTS 510
QY 350 KDSKQPR-----KKEGRYRQYDHE 368
Db 511 KONKKEKRIQIDVDEKTSKNFRSIFYDQE 538
```

```
RESULT 15
US-10-011-588-45
; Sequence 45, Application US/10011588
; Publication No. US20020168727A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Leonard
; TITLE OF INVENTION: RECOMBINANT LIGHT CHAINS OF BOTULINUM
; TITLE OF INVENTION: NEUROTOXINS AND LIGHT CHAIN FUSION PROTEINS FOR USE IN
; TITLE OF INVENTION: RESEARCH AND CLINICAL THERAPY
; FILE REFERENCE: A34796 067252.0113
; CURRENT APPLICATION NUMBER: US/10/011,588
; CURRENT FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: 09/910,186
; PRIOR FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: 09/611,419
; PRIOR FILING DATE: 2000-07-06
; PRIOR APPLICATION NUMBER: 60/246,744
; PRIOR FILING DATE: 2000-11-06
; PRIOR APPLICATION NUMBER: 60/311,966
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 45
; LENGTH: 848
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant protein encoded by SEQ ID NO:44
US-10-011-588-45

Query Match      5.9%; Score 132.5; DB 13; Length 848;
Best Local Similarity 20.4%; Pred. No. 0.18;
Matches 79; Conservative 64; Mismatches 129; Indels 115; Gaps 16;

QY 25 GSIGPSIVCKSIQNAEQSLQEEQGLDLTVNRMOEQNTQOGDGVLDLSTKTSIKSE 84
DB 273 GGHDPSPVISTDMNIYKALQNFQD-----IANRLNVSSAOGSGI-DISLYKQIYKKN 326
QY 85 ESSICDPSSNSVAGRLHRNREDYVERSAEFADGLSKALKDIQSGALDINKAGILYGI 144
DB 327 YDFVEDPNGKVS-----DKDKF-----DKLYKALMFGFTETNLAG-EYGI- 366
QY 145 QKILLHL-EALP-----AGKPAPKPKTRDFH-----DSYSYKDSKE 181
DB 367 -KTRYSYFSEYLPPIKTEKLLDNITVTQNGFNIAKSLKTEFNGQNKAVKAEYEISL 425
QY 182 TCALQKVALWARAQARTSKSLNLLSETSEIKFPPTASTYLHQLTLQKQWVTFPKKNESL 241
DB 426 EHLVIYRIAMCKPVMYKNTGKSEQCIIVNNEDLFFIAN-----KDSFSKDLAKAETI 477
QY 242 QYETSNPTVQ-----LQIP-QLRVSVSVKSQP 267
DB 478 AYNTQNTNTIENFSIDQLILDNLSSGIDLFPNTEPTNFDIDIDIEVYIKQSALKKIFV 537
QY 268 DGSGLLDVMYQVSKTSSVLEGSALQKLKNIPLKPKK-----IECSGFTVTHSSVDSY 318
DB 538 DGDSLFEYLHAQTFFSNI-ENQLTNSLNDALRNKKVYTFPTNLNVEKANTVVGAS----- 592
QY 319 FLHGDLSPLCLNSKNGTVDG-TSENTE 344
DB 593 -----LFVNVKVGVIDDFTSESTQ 611

Search completed: February 24, 2005, 02:06:50
Job time : 132 secs
```



Result No.	Query			DB	ID	Description
	Score	Match	length			
1	183	8.1	185	2	T24276	hypothetical prote
2	133.5	5.9	1109	2	A40801	phosphoprotein pho
3	131.5	5.8	997	2	T43523	cuti7 protein - fi
4	131	5.8	1297	2	S39791	neurotoxin - Clost
5	129.5	5.8	545	2	T19172	hypothetical prote
6	129.5	5.8	1560	2	T42727	proliferation prote
7	127	5.6	534	2	S61185	hypothetical prote
8	126.5	5.6	983	2	B49284	immediate-early pr
9	126.5	5.6	1078	2	T44332	hypothetical prote
10	126	5.6	433	2	T50395	actin-related prote
11	125.5	5.6	1819	2	A71928	cag island protein
12	125	5.6	948	2	T41496	conserved hypothet
13	124.5	5.5	2954	2	T14156	kinesin-related pr
14	124.5	5.5	3187	2	JC5837	364K Golgi complex
15	123.5	5.5	1827	2	T16270	hypothetical prote
16	123	5.5	1017	2	PC4035	cell-cycle-depende
17	123	5.5	1269	2	F84730	probable myosin he
18	122.5	5.4	958	2	T20621	hypothetical prote
19	122	5.4	911	2	S51441	hypothetical prote
20	122	5.4	1164	2	S46769	hypothetical prote
21	122	5.4	1435	1	BVBYL1	guanine nucleotide
22	121.5	5.4	1927	2	G64585	cag pathogenesis
23	121.5	5.4	2938	2	T30249	cell proliferation
24	121	5.4	833	2	T43446	hypothetical prote
25	120	5.3	1170	2	A56157	chromosome segrega
26	119	5.3	1148	2	A49651	replication factor
27	118.5	5.3	1313	2	F96673	hypothetical prote
28	118.5	5.3	2588	2	T14342	NSD1 protein - mou
29	118.5	5.3	5105	2	T32650	hypothetical prote

A: Molecule type: tRNA

A;Cross-references: UNIPROT:Q22051; EMBL:Z68010; PIDN:CAA92009.1; GSPDB:GN00028; CESP:T01

A;Residues: 1-1109 <TAN>  
A;Cross-references: UNIPROT:Q00756; GB:M65109; NID:gl65692; PIDN:AAA31462.1; PID:gl65693  
R;Dent, P.; Campbell, D.G.; Hubbard, M.J.; Cohen, P.  
FEBS Lett. 248, 67-72, 1989

A;Title: Multisite phosphorylation of the glycogen-binding subunit of protein phosphatase  
A;Reference number: S04004; MUID:89252053; PMID:2542090

A;Accession: S04004

A;Status: preliminary

A;Molecule type: protein

A;Residues: 33-68 <DEN>

C;Keywords: phosphoprotein; phosphoric monoester hydrolase

Query Match 5.9%; Score 133.5; DB 2; Length 1109;  
Best Local Similarity 22.9%; Pred. No. 1.6; Indels 83; Gaps 17;  
Matches 89; Conservative 53; Mismatches 163

QY 7 QFAIEYISGKTKQENRNGSIQPSIVCKSIQMNQAEISLQEEQEGPLDITVNRMOEQNTQ 66

Db 208 EFCIRYETSVGTFSWNNGT-NYTLVCQ-----KKEPEPFGKPL-----BEAPSK 252

QY 67 QDGVLDLSTKTKTSIKSESSICDPSESSVAGRLHNRNEDYVER---SAEFADGLLS-- 121

Db 253 QKKGCLKVKSSK-----ESS--ETSENNFEN--SKIADTVIPTIVCSHEEKEDLKSSY 303

QY 122 KALKDIQSGALDINKAGILYIPQKTLHLHLALPAGK-----PASFKNKTDRDFHDSYS 175

Db 304 QNVKDVNTEHDEHNEKELEMLNQRIRTRCAASEYKGNLTSLSDPSNIKNKPEELQKNQS 363

QY 176 YKDSKETCAVLQKVALWARAQARTKSKNLLLETSEIKFPTASTYHLQITLQKMTQPK 235

Db 364 H---SEACTDLSQRLSPGSSAESLKGDFYHTE---KYSSGNESSHQ-----PSDMG 410

QY 236 EKNESIQYETSNPTVQKLPQARVSSVSKSQPDGSLDV-----MYQVSKTSSVLEGS 289

Db 411 EINPSLGGTSDGVSQVHLHSSKEILDNDNANPAHGSGRGEISCFPGQLKASNLKKYEGG 470

QY 290 ALQKLKNILPKQNKIECSGVPVTHSSVDFLHGLDLSPLCLNSKNGVTVDGTSNTDGLDR 349

Db 471 AENS-----EMKDCLEPRDVLHKASDYF-----KKSTENRPSE--EDYGTGS 510

QY 350 KDSKQPR-----KKGGRYQYDHE 368

Db 511 KDKKRIQLDVDEKTSKNPRSFYDQE 538

## RESULT 3

T43523

cut17 protein - fission yeast (Schizosaccharomyces pombe)

C;Species: Schizosaccharomyces pombe

C;Date: 21-Jan-2000 #sequence revision 21-Jan-2000 #text\_change 09-Jul-2004

C;Accession: T43523; T41649, T41700

R;Morishita, J.; Matsusaka, T.; Yanagida, M.

submitted to the EMBL Data Library, August 1999

A;Description: Fission yeast cut17 is required for chromosome segregation.

A;Reference number: Z22536

A;Accession: T43523

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-997 <HAR>

A;Cross-references: UNIPROT:O14064; EMBL:AB031034; PIDN:BAAB3415.1

R;Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.

submitted to the EMBL Data Library, August 1998

A;Reference number: Z22007

A;Accession: T41649

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-997 <HAR>

A;Cross-references: EMBL:AL031323; PIDN:CAA20434.1; GSPDB:GN00068; SPDB:SPCC362.02c

A;Experimental source: strain 972h-; cosmid c962

R;Wedler, H.; Duysterhoef, A.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.

submitted to the EMBL Data Library, October 1999

A;Reference number: Z22010

A;Accession: T41700

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 932-997 <WED>

A;Cross-references: EMBL:AL121859; PIDN:CAB58376.1; GSPDB:GN00068; SPDB:SPCP31B10.10c

C;Genetics:

A;Gene: cut17; SPCC962.02c; SPDB:SPCP31B10.10c

A;Map position: 3L

A;Introns: 43/3

Query Match 5.8%; Score 131.5; DB 2; Length 997;

Best Local Similarity 21.7%; Pred. No. 1.8; Indels 141; Gaps 24;

Matches 109; Conservative 77; Mismatches 175

QY 30 STVCKSIQMNQAEISLQEEQ-----EGPLDITVNRMOEQNTQDGVLDLSTKTKTS 80

Db 408 SVVSKSKSISSVSVSGKEQHTEKQVALETPEQKQKVEDEHLNLQG-SFIEESTKQ-P 465

QY 81 TKSESSICDPSESSVAGRLHNRNEDYVERSAEFADGLLSK-----ALKD 126

Db 466 ISSKPESTSPDMTAAATGGRV-----SSSFRDKILQTNFSPRSTIDSFNSIKKR 516

QY 127 IQSGALDIN-KAGILYIPQKTLHLHLALPAGKPAFPKN-----KTRDFHDSYSYKDSKE 181

Db 517 NSEANDENETNLKIPKPKKQFQ-EVLQS-----XNLLVSTSDSHEPVPKVTEDSQ 569

QY 182 TCAPVLQKV-----ALWARAQARTKSKNLLLETSEIKFPTASTYHLQITLQK 229

Db 570 TAIHVSKPFEDLENKSESEQSLQLESSENDKPLIDILPLAIAK-----RKDN 618

QY 230 MYTQPEKNES-----LQYETSNPTVQKIPQARVSSV-----SKSQPDGSGLLDVMYQ- 278

Db 619 LVSGVLEKGTSTSKTKFDTSDIVDF-LEKPKTEISEVLPEEKRAICADESQTVRVSDIR 677

QY 279 -VSKTSSVLEGSALQKLNILPKQNKIECSGVPV-----HSSVD-----SY-- 318

Db 678 GVTKTRDVSSPVSDKSENV-----NHEANSCHTVMVHSSLDPPQIVQPNLESGSYLK 733

QY 319 -----FLHGLDLSPLCLNSKNG-TVDGTSNTDGLDRKDSKQPKKGRYR 363

Db 734 DLPDRNVGSEKVTFTQEDDINSPLQSKNNQTVAVNTETSDKLOEKA----- 782

QY 364 QYDHEI-----MEALAMVMSCKMSVSKAQGIYGVPHSTLEYKVKERSGTLKTPPKK-KL 417

Db 783 --NHELENIEKTEEKLTEV--DKVSLSDAPPQEIKNRSTSVQNGTRSVSKNTPPEKTKV 838

QY 418 RLPTDGLYNNMTSDGTGCKNSS 439

Db 839 DKIDNVSKKDVETSPGSCETSS 860

## RESULT 4

S39791

neurotoxin - Clostridium botulinum

C;Species: Clostridium botulinum

C;Date: 07-Oct-1994 #sequence\_revision 01-Dec-1995 #text\_change 16-Jul-1999

C;Accession: S39791

R;Campbell, K.; Collins, M.D.; East, A.K.

Biochim. Biophys. Acta 1216, 487-491, 1993

A;Title: Nucleotide sequence of the gene coding for Clostridium botulinum (Clostridium a

A;Reference number: S39791; MUID:94032745; PMID:8268233

A;Accession: S39791

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1297 <CAM>

A;Cross-references: EMBL:X74162; NID:G441275; PIDN:CAAS2275.1; PID:G441276

C;Superfamily: tetanus toxin

C;Keywords: neurotoxin

Query Match 5.8%; Score 131; DB 2; Length 1297;

Best Local Similarity 20.2%; Pred. No. 2.7;

Matches 78; Conservative 64; Mismatches 131; Indels 114; Gaps 15;

QY 25 GSIGPSIVCKSIQMNQAEISLQEEQEGPLDITVNRMOEQNTQDGVLDLSTKTKTSIKSE 84

```
Db 273 GGHDPSSVSTDMYINKALQFQD-----IANRLNIVSAQSGI-DISLYKQYKXK 326
QY 85 ESSICDPSSNSVAGRLHRNREDYVERSABFADGLLSKALKQIQSGALDINKAGILYGP 144
Db 327 YDFVEDPNGKYSV-----DKKXF-----DKLYKALMFGFTETNLAG-EVGI- 366
QY 145 OKTLLHLH-EALP-----AGKPASFNKTRDFH-----DSVSYKDSKE 181
Db 367 -KTRYSYFSEYLPPIKTEKLDNTIYQNEGFNIASKNLKTEFGNQKAVNKEAYEISL 425
QY 182 TCALVOKVALWARAQARTSKNLNLETSEIKFPTASTYLHQLTIQKMWTFQKKNESL 241
Db 426 EHLVIVRIACKPMYKNTGKSEQCIIVNNEDLFFIAN-----KDSFKDLAKAEYI 477
QY 242 QYETSNTPTQV-----LQIP-QLRVSVSVSKSQP 267
Db 478 AYNTQNTIENFSDIQILINDLSSGIDLPLNTEPFTNFDDIDIPVYIKQALAKKIFV 537
QY 268 DSGGLDVMYQVSKTSVLEGSALQKLNILPKONK-----IECSGPVTHSSVDSY 318
Db 538 DGDLSFELYLHAQTFPSNIENLQITNSLNDALRNNKVTYTFPSTNLVEKANTVVGAS----- 593
QY 319 FLHGDLSPCLNSKNGTVDG-TSENTE 344
Db 594 -----LFVNVWKGVIDDFTSESTQ 612

RESULT 5
T19172
hypotheical protein F18C12.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T19172; T21088
R;Harris, B.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19083
A;Accession: T19172
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-545 <W1>
A;Cross-references: UNIPROT:O17584; EMBL:Z81466; PIDN:CA803870.1; GSPDB:GN00019; CESP.F1
A;Experimental source: clone C09H6
R;Harris, B.
submitted to the EMBL Data Library, June 1996
A;Reference number: Z19371
A;Accession: T21088
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-545 <W12>
A;Cross-references: EMBL:Z75536; PIDN:CAA99833.1; GSPDB:GN00019; CESP.F18C12.3
A;Experimental source: clone F18C12
C;Genetics:
A;Gene: CESP.F18C12.3
A;Map position: 1
A;Introns: 171/3; 222/2; 316/3; 368/3; 409/1; 409/3; 493/1

Query Match 5.8%; Score 129.5; DB 2; Length 545;
Best Local Similarity 20.8%; Pred. No. 1.1;
Matches 93; Conservative 72; Mismatches 174; Indels 109; Gaps 20;

QY 2 KKMIRFA-----IEYISGKTQNRNGSIGPSIVCKSIQMNQNSIQE---RQEGP 52
Db 175 KDSVAKFAALNTKSLAELKINGKTEVAER--LMYKLVCSIIIVEKIQ-SIREKYTEMNKT 231
QY 53 LDLTVMRMOBQNTQQGDVLDLTKTKTSI---KSESSICDPSSNSVAGRLHRNREDYV 109
Db 232 AAFPQRLGOQYAKPEKSKSKTEKELNIENLEDNNEDESENKEEIEENEEDYD 291
QY 110 ERSABFADGLLSKALKDIQSGALDINKAGILYGPQKTLHLHLEAL-----PAGKPASF 163
Db 292 QSDIEMLDS-----DEBEAGEAAKNRRNLLGLIGVQEDKNRPSLAPKR 337
```

```
QY 164 KXKTRDPHDSYKDSKETCAVLQKVALWARAQARTSKNLNLETSEIKPPTASTYLH 223
Db 338 KLVEEDEDVEVTYKQKMKQIFKE-----LKSKEVKNP----- 372
QY 224 QLTQKMWTFQKKNESLOVETSNPTVQLKIPOLRV--SSVSKSQDPGSGLLDMYQVSK 281
Db 373 -----KSNKSTKKPKTSAPIVKVKEKVEEVEVSDDESQKTLV-MKVDLSK 420
QY 282 TSVLEGSALQKLNILPKONIECSGPVTHSSVD-----SYPL--HGDLSPCLNSKNG 334
Db 421 GGIKAKA---QKFTTAPKSAKI--VAPVSEDDDDSSFFLPKSGVVPAPRKIIIPKK- 474
QY 335 TVDGTSENTEGLDRKDSKQPKRGYRQYDHEIMEEALAMVMSCKMSVSKAQGIYGPV 394
Db 475 -----PSENEK-VDRKFQKGTQKK-----SEAVVEKKKG-----SSKSAVSGEM 513
QY 395 H-STLEYKVKERSGTLKTPPKKLRPD 421
Db 514 HPSWIASQLKKELASAKPCGKKITFGD 541

RESULT 6
T42727
proliferation potential-related protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004
C;Accession: T42727
R;Witte, M.M.; Scott, R.E.
submitted to the EMBL Data Library, November 1998
A;Reference number: Z22246
A;Accession: T42727
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-1560 <WIT>
A;Cross-references: UNIPROT:P97868; EMBL:U831913; NID:g3858884; PID:g3858885; PIDN:AAC7243
A;Experimental source: strain Balb/C
C;Genetics:
A;Gene: P2P-R
C;Function:
A;Description: involved in hnRNP association and Rb1 binding
F;57-107/Domain: RING finger homology <RRN>

Query Match 5.8%; Score 129.5; DB 2; Length 1560;
Best Local Similarity 20.5%; Pred. No. 4.4;
Matches 103; Conservative 81; Mismatches 180; Indels 139; Gaps 23;

QY 15 KSGKTOENRNGSIGPSIVCKSIQMNQNSIQEQQGGLDLT-----VNRMOEQNTQQG 68
Db 786 KSDKTKRKSDG-----SATAKKNVLKPSGPKQPKVDGDKREKSPRSPPLKAKKEATK-- 839
QY 69 DGVLDLSTKTKTSIKSESSICDPSSNSVAGRLHRN---REDYVRSABFADGLLSKAL 124
Db 840 ---IDSVPKSSSSQKDEKVTGTPRKAHSAKSAHEQAKPAKQKVKDC-----SKDI 899
QY 125 KDIQSGALDINKAGILYGPQKTLHLHLEALPAGKPASFNKTRDPHDSYKDSKETCA 184
Db 890 KSEKPAKD-EKA---KKPEKNKLIDSKGKKRKRKTEESKVDKDF-ESSSMKISKVEGT 943
QY 185 VLQKVALWARAQ-----ERT-EKSKNLNLETSEIKFPTASTYLHQLTIQK--VTQ 233
Db 944 EIVQSPKMKMGDVEKLETPKDKIASSTT-----PAKKTKLNRETGKKIGNAST 998
QY 234 FKKNESLOVETSNPTVQLKIPOLRVSSVSKSQPDGS-GLLDVMYQVSKTSVLEGSALQ 292
Db 999 TKPEKLESTSS-----KIQEKVKGAKRVAGSESSSTLVDTYTSST-GGSPVR 1051
QY 293 KUK-----NILP-----KONKIECSGPVT 311
Db 1052 KSEKTDTRTVIKTMEEYNNNDTAPEDVIMIHVPQSKWDKDDFESEEDVKTTPQIQ 1111
QY 312 HSVDYSFLHGDLSPCLNSKNGTV--DGTSENTEGLDRKDSKQPKRGYRQYDHEI 369
Db 1112 SVGKPSII-----KNVTKPSATAKYTE-----KSEQPEKQLQKPEASHL 1155
```

QY 370 MEEATAMWSGKMSVSKAOGIYGVPHSTLEYKVKERSGTLKTPPKKL----- 417  
Db 1156 MQHEL---RSSKGSASSEK-----RAKDREHSGSEKDNPKRKSQAQPKSESTD 1203  
QY 418 RLPTDGLYNWTDGSGCKNSK 440  
Db 1204 RLSEQHFKTLQSSEKETSSEK 1226  
RESULT 7  
S61185  
hypothetical protein YDR299w - yeast (Saccharomyces cerevisiae)  
N/Alternate names: hypothetical protein D9740.7  
C/Species: Saccharomyces cerevisiae  
C/Date: 23-Feb-1996 #sequence\_revision 01-Mar-1996 #text\_change 09-Jul-2004  
C/Accession: S61185  
R/Ding, H.  
submitted to the EMBL Data Library, June 1995  
A/Description: The sequence of S. cerevisiae cosmid 9740.  
A/Reference number: S61160  
A/Accession: S61185  
A/Molecule type: DNA  
A/Residues: 1-534 <DIN>  
A/Cross-references: UNIPROT:Q06631; EMBL:U28374; NID:g849214; GSPDB:GN00004  
C/Genetics:  
A/Gene: SGP:BPR2; MIPS:YDR299w  
A/Cross-references: SGD:S0002707  
A/Map position: 4R

Query Match 5.6%; Score 127; DB 2; Length 534;  
Best Local Similarity 21.4%; Pred. No. 1.5;  
Matches 92; Conservative 61; Mismatches 157; Indels 120; Gaps 16;  
QY 5 IROFAIYISKSGKTONRNGSI-----GPSIVCKSQMNOAENSLQEQEGPLDLTV 57  
Db 9 ISDIAIKPVNKDFDIEENASLFOHNEKNES-----DLSYGNSTTEETKKAHYLEV 62  
QY 58 NRWQEQNTQOGDGLDLSTKK-TSIKSESSICDPSSSENSVAGRLHRNREDYVERSAEF- 115  
Db 63 -----EKSRLRAEKGLENDPKYTGKSGRQALVEESENDEDEEEEEEKEEDALSFR 118  
QY 116 -----ADG-----LSSKAL-----KDQSGALDINKAGI 139  
Db 119 TDSDEVEIDEESDAGGTEEAQKRUHLSKLIQKQETKQAINKLSQSVQRDASKG-- 176  
QY 140 LYGIPOKTL-----LHLEALPAGKPKASFKNTRDFHDSYKQMSKETCAVLQKVALWA 193  
Db 177 -YSILOQTKLFDNIDLRIKLOKAVIAANKPLPTTESWEAKMDDEETKRLUK----- 229  
QY 194 RQAARTEKSKLNLETSEIKF-----PTASTYLHQLTLQKMTQFKEKNESLOVETS 246  
Db 230 --ENEKLFNLFNRLINFRKLFQGDHITQNEEVAKHLSKRSLSKELYOETNSLSELK 287  
QY 247 N-PTVQLKIPLQRYSSVSKSQPDGSL-----LDVMYQVSKTSVLEGGALQKLN 296  
Db 288 EYRTAVLKNKSTKVSSASGNAALSNKFKALNPADYQVENQLSDMSRLMKRTKLN-RN 346  
QY 297 ILPKQNKIECS-----GPVTHSSVDSYFLHGLDPLCLNSKNGTVDGTSNTDGLD 348  
Db 347 ITPLYFOKDCANGRLPELISPVVKDSVDD-----NENSDGLD 384  
QY 349 RKDSKQPRKK 358  
Db 385 IPKNYDFPKK 394

RESULT 8  
B49284  
immediate-early protein RF3/RF4 - human herpesvirus 6 (strain Z29) (fragment)  
C/Species: human herpesvirus 6  
C/Date: 01-Dec-1995 #sequence\_revision 01-Dec-1995 #text\_change 09-Jul-2004  
C/Accession: B49284

R/Chou, S.; Marousek, G.I.  
Virology 198, 370-376, 1994  
A/Title: Analysis of interstrain variation in a putative immediate-early region of human  
A/Reference number: A49284; MUID:94082474; PMID:8259673  
A/Accession: B49284  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-983 <CHO>  
A/Cross-references: UNIPROT:Q69530; GB:L21760; NID:g347260; PIDN:AAA16547.1; PID:g347261

Query Match 5.6%; Score 126.5; DB 2; Length 983;  
Best Local Similarity 18.7%; Pred. No. 3.6;  
Matches 73; Conservative 73; Mismatches 164; Indels 81; Gaps 13;  
QY 9 AIEYISKSGK-TOENRNGSIGSVCKSIQMNQAEENSLQEQEGPLDLTVNRMQEONTQQ 67  
Db 642 AVSQKCKSKRTAKRKNVPKPS-KSKKIKLRLPET-----TNVIVISSEDEED 692  
QY 68 GGVLDLSTKTSIKSESSICDPSSSENSVAGRLHRNREDYVERSAEFADGLLSKALKDI 127  
Db 693 GNNIIDKSMLEKTIKSE-----PNSESSSDCTSEDNYLH-----LSDYDKVI 737  
QY 128 QSGALDINKAGILYGIPOKTLHLEALPAGKPKASFKNTRDFHDSYKQMSKETCAVLQ 187  
Db 738 NNGHCQSK-----GFPSPVTPIRSMFG-----THDIRNKF-----VPK 772  
QY 188 KVALWARAQARTE-----KSKLNLETSEIKFTASTYLHQLTLQKMTQFKEKNESL 241  
Db 773 KHLWFMKTHKVDNVCVHSHSAKMNKNDSDVTEANHCFINHFVPIKTDDEYEKENVSY 832  
QY 242 QY-----ETSNPTVOLKIPOLRYSSVSKSQPDGSLDLVMYQVSKTSVLEGGAL 291  
Db 833 TYSKIEDSKTDLDTPTPKLITEMVMENFMDLTDIIKHGIANHCQDLSKVTYTHTAC 892  
QY 292 OKLKNILPKQNKIECSGPVTHSSVDSYFLHGLDPL-----CLNSKNGTVDGTSNT 343  
Db 893 EKNLVANSQNLVTAETQIFDPQGT-----GNNSPILAIINDTTQCDENRCTGTSNDN 947  
QY 344 EDGLDRKDSKQPRKKGRYQY--DHEIMBE 372  
Db 948 EKCTIRSDCNSDKMEVFKLDGYPSDYDPFEE 978

RESULT 9  
T44232  
hypothetical protein U90 [imported] - human herpesvirus 6 (strain Z29)  
C/Species: human herpesvirus 6  
A/Variety: strain Z29  
C/Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 09-Jul-2004  
C/Accession: T44232  
R/Dominguez, G.; Dambaugh, T.R.; Stamey, F.R.; Dewhurst, S.; Inoue, N.; Pellett, P.E.  
J. Virol. 73, 8040-8052, 1999  
A/Title: Human herpesvirus 6B genome sequence: coding content and comparison with human  
A/Reference number: Z22734; MUID:99412318; PMID:10482553  
A/Accession: T44232  
A/Status: preliminary; translated from GB/EMBL/DDB  
A/Molecule type: DNA  
A/Residues: 1-1078 <DOM>  
A/Cross-references: UNIPROT:Q9QJ15; EMBL:AF157706; PIDN:AAA49675.1  
A/Experimental source: strain Z29; variant B  
C/Genetics:  
A/Introns: 32/2; 103/1  
A/Note: U90

Query Match 5.6%; Score 126.5; DB 2; Length 1078;  
Best Local Similarity 18.7%; Pred. No. 4.1;  
Matches 73; Conservative 73; Mismatches 164; Indels 81; Gaps 13;  
QY 9 AIEYISKSGK-TOENRNGSIGSVCKSIQMNQAEENSLQEQEGPLDLTVNRMQEONTQQ 67  
Db 737 AVSQKCKSKRTAKRKNVPKPS-KSKKIKLRLPET-----TNVIVISSEDEED 787  
QY 68 GGVLDLSTKTSIKSESSICDPSSSENSVAGRLHRNREDYVERSAEFADGLLSKALKDI 127

```
Db 788 GNNIIDKSMLEKTIKSE-----PNSSESSESDCTSEDNYLH-----LSDYDKVI 832
Qy 128 QSGALDINKAGILYGIPOKTLHLLEALPAGKPAFPKNTTRDPHDSYVSKYDSEKTCVILQ 187
Db 833 NNGHCQSK-----GFPSPVFTPIRSMGP-----THDIRNKF-----VPK 867
Qy 188 KVALWARAQARTE-----KSKLNLETSEIKFPASTYVHLQTLTKQMTQPKENESL 241
Db 868 KHLWLMFMRKTHKVDNCVHSSAKMKNKDSVDVTEANHCFTINHFVPKTDDEEYERENVSY 927
Qy 242 QY-----ETSNTPTQKIPQLRVSVSQSPDGSGLLDVMYQVSKTSSVLEGSAL 291
Db 928 TYKIQDSKTDLEIPTTKKLIITEMWENFMJLTDIIKIGIAKQDLASKYTVITHAC 987
Qy 292 QKLNILPKQNKIECSGPVTHSSVDSYFLHGLDLSPL-----CLNSKNGTVDGTSNT 343
Db 988 EKLNANVSQNLVTAETQIFDPQGT-----GNNSPILNIINDTTCQDNCRCTEGTSNDN 1042
Qy 344 EDGLDRKDSQPKRKRGRYQY--DHEIMEE 372
Db 1043 EKCTIRSDCNDRKMEVFKLDGYPDYDPPE 1073

RESULT 10
T50395
actin-related protein [imported] - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jul-2004
R;Beck, A.; Borzym, K.; Reinhardt, R.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, January 1999
A;Reference number: Z25067
A;Accession: T50395
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-433 <BEC>
A;Cross-references: UNIPROT:Q9P7X7; EMBL:AL136535; PIDN:CAR66436.1; GSPDB:GN00067; SPDB:
A;Experimental source: strain 972h(-); clone pl p23A10
C;Genetics:
A;Gene: SPDB:SPBP23A10.08
A;Map position: 2
C;Superfamily: actin

Query Match 5.6%; Score 126; DB 2; Length 433;
Best Local Similarity 22.8%; Pred. No. 1.3;
Matches 95; Conservative 56; Mismatches 152; Indels 114; Gaps 19;

Qy 6 RQFAIEYISKGTQENRNGSIGPSIVCKSIQMNQAE-----SLQEB-QSGPLD 54
Db 49 RLFGEYIYKSNPGMEIKN-----AIRNGWVENMDVTVDLWRYGLEQQLKTNPLE 98
Qy 55 LTVNRMQEQNTQQGDVLDLSTKTSIK-----SSESSIC--DPSSENS 96
Db 99 HPILITEPDPNPNRVTLETWFESURCPATYLAKQETCAAFASGKGTACLVDIGAERS 158
Qy 97 VAGRLHRNREDYVERSAEAFGLLSKALDKIQSGAL--DINKAGILYGIPOKTLHLLEAL 155
Db 159 SVSAIY---DGFVLQGYQVQHSGNAINDILAQTLRDKN-----FEVWPKYLVSKNPV 210
Qy 156 PAKGPASFRNKRTRDFHDSYS-----YKDSKETCAVLQKVALWARAQARTEKSKNLNL 208
Db 211 EIQGPANCELPRDITDSYHQFQVRVYDEWKECALISDVPF-----SSETTI 259
Qy 209 ETSEIKFPASTYL-----HOLTQKMTQFKEKNESLOYETSNPTVQ----- 251
Db 260 AESEFEFPDGRNMFGAERYQIPEHLFV-----PGSDEEMNEEPSKPIEQTENNEVSQODSS 316
Qy 252 -----LKIPLRVSSVSQSPD--GSGLLDVMYQVSKTSSVLEGSALQKLNILPK- 300
Db 317 VNTTSRILGIPQLFQNCITSECVDVIRASLLNNVI--VCGGTSIMQGFSL-RLQNELSKL 373
Qy 301 ----QNKIECSGPVTHSSVDSYFLHGLDLSPLCLNSKNGTVD-----GTSNTEDGLDR 349
```

```
Db 374 YPGSRUKIHASGHVRSYASWLGSGILSSL-----GTFHQLWISRQRYEHEGSDR 424

RESULT 11
A71928
cag island protein - Helicobacter pylori (strain J99)
C;Species: Helicobacter pylori
A;Variety: strain J99
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004
R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric patho-
A;Reference number: A71800; MUID:99120557; PMID:99233682
A;Accession: A71928
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1819 <ARN>
A;Cross-references: UNIPROT:Q9ZLV0; GB:AE001481; GB:AE001439; NID:g4155005; PIDN:AAD0604
A;Experimental source: strain J99
C;Genetics:
A;Gene: orf13/14

Query Match 5.6%; Score 125.5; DB 2; Length 1819;
Best Local Similarity 19.7%; Pred. No. 9.5;
Matches 102; Conservative 70; Mismatches 180; Indels 165; Gaps 18;

Qy 14 SKSGKTOENRNGSIGPSIVCKSIQMNQ-----AENSLOEQEGPLDLTVNRMQEQNTQ 67
Db 72 SGNETSSESSNGSLADKLFKARKLVDDKRPTQKSLDEETQ---KLNEDDQENNEHQ 128
Qy 68 GDGVLDLSTKTSIKSEESSICDPSSENSVAGRLHRNREDYVERSAEAFGLLSKALKDI 127
Db 129 EETQTDLIDDETSEKTSQDQSPDLSNEATEA--NHFDLLKESTESSDNHLN----- 180
Qy 128 QSGALDINKAGILYGIPOKTLHLLEALPAGKPAFPK-----TRDFHDSYSYKD 178
Db 181 -----PTESSDNHLNPTETKTQETKTHFDEDKPREITDSDNDQBIIG 224
Qy 179 SKE-----TCAVLQKVALWARA----- 195
Db 225 SKKKYIIGVAVLVILFIRSIFHYFVPLEDKSSRSKORNLNVNDEIQRBYNRL 284
Qy 196 QAERTKSKL-----NLLETSEI--KFPTASTY-----LHQL 225
Db 285 LKERNEKNMIDKNLFPNDPNRTLNYLNIAIEIDKNPLAFYECISNGGNVEECLKLI 344
Qy 226 TLQKMTQPKENESLOYETSNPTVOLKTPQLRVSVSKSQPDGSGLLDVMYQVSKTSSV 285
Db 345 KDKKLQDQMKLTLEAYNDCIKN-----AKTEBERIKCLDLIKDENLKKSL 389
Qy 286 LSGS-----ALQKLNILPKQNKIECSGPVTHSSVDSYF-----LHGDLSPL--CLNS--- 331
Db 390 LNQKQVQVALDCLNKAKTDEERKECLKLINDPEIRKFKLELQKLEQEKYKDCIKNAKT 449
Qy 332 ---XNGTVDGTSNTEDGLDR-----KDSKQPKRKRGRYQYVDHEIMEEATA--MMSG 380
Db 450 EAEKNECLKGLEAERLKKQALDCLNKAKTDEERKECLKNIPDLQKELLADMVKAY 509
Qy 381 KMSVSKAQGIYGVPHSTLEYKVKERSGTLKTPPKKL 417
Db 510 KDCVSRAR-----NEKEQCEKLLTPPAKKL 536

RESULT 12
T41496
conserved hypothetical protein SPPC622.16c - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
R;Seeger, K.; Harris, D.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.
```

submitted to the EMBL Data Library, October 1998

A:Reference number: Z21998  
A:Accession: T14156  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-948 <SEE>  
A:Cross-references: UNIPROT:O94603; EMBL:AL033127; PIDN:CAA21872.1; GSPDB:GN00068; SPDB:  
A:Experimental source: strain 972h-; cosmid c622  
C:Genetics:  
A:Gene: SPDB:SPCC622.16c  
A:Map position: 3

Query Match 5.6%; Score 125; DB 2; Length 948;  
Best Local Similarity 19.6%; Pred. No. 4.3;  
Matches 93; Conservative 74; Mismatches 198; Indels 110; Gaps 19;  
QY 7 QFAEYISKSGK--TQENRNGSIGPSIVCKSIQMNQ-----AENSLOEQE--GPLD 54  
DB 514 EICVDVQKFGAMITVHRRSAKHPSCNCFSLQTLIDSGPKPANNYSQHQSNTFIGVVI 573  
QY 55 LT-----VNRMOEQNTQGGDGLD-----STKKT-----SIKSESSICDPSS 94  
DB 574 STNHNIIKKQESQIQTKNNCSFQLVKRIKSTKKAWSRIKAFKRENTRCNPLSS 633  
QY 95 NSVAGLHRN--REDVVERSAEPADGLSKALXDIQSGALDINKAGILYGIPOKTLHL 152  
DB 634 -----LHATTFREDIVRPKINSFVLEQLIFALSFALNTWTFSPFLNHSFNENIALSK 687  
QY 153 EALPAGKASFKNTRDFHDSY---SYKSKETCAVLQKVALWARAQAERTKSKLNLE 209  
DB 688 ETENFGEANCENTDTLTFTWGDGFRPSDSIC-----YNDENLLE 729  
QY 210 TSEIKPTASTYLHOLTLQKMTQFKEKNELOYESNPTVOLKIPOLRVSSVSKOPDG 269  
DB 730 TAN--SDAEASTHELEQLP-NAVNEREVDISQDTPSTAL---DTRVTRKVDLSPEF 782  
QY 270 SGLL-----DMVYQVSKTSVLEGSALQKLNILPKONKIEGSPVTHSSVDSYFLHGD 323  
DB 783 SNILSPSSNDSSFQDLDDLSPPSSNLKQIQKVPV-QNSLEFSVGEKKAAYSLHT 841  
QY 324 LSPCLNSKNGTVGTSNTEDGLDRKDSQPKPKGRVQYDHEIMEEAIAVMWSG--- 380  
DB 842 FSKRLSMENKPDPTTKVPLKYNIOHEEMKAYRRKN-----DLEIDQHPASSKSGISN 895  
QY 381 -----KMSVSKAGGIYGVPHSTLEYKVKERSGTLPKPKKLRPLDPTGLYNTD 429  
DB 896 GRNNKEVNLTKAENV-GI-----KRRIMKNENIYDFED 930

RESULT 13  
T14156  
kinesin-related protein - African clawed frog  
C:Species: Xenopus laevis (African clawed frog)  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: T14156  
R:Wood, K.W.; Sakowicz, R.; Goldstein, L.S.; Cleveland, D.W.  
Cell 91, 357-366, 1997  
A:Title: CENP-E is a plus end-directed kinetochore motor required for metaphase chromosome  
A:Reference number: Z17893; MUID:98028574; PMID:9363944  
A:Accession: T14156  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-2954 <WOO>  
A:Cross-references: UNIPROT:O42263; EMBL:AF027728; PIDN:G2586070; PID:G2586071; PIDN:AA06  
C:Genetics:  
A:Gene: XCBP-E  
C:Superfamily: centromere protein E; kinesin motor domain homology

Query Match 5.5%; Score 124.5; DB 2; Length 2954;  
Best Local Similarity 18.5%; Pred. No. 21;  
Matches 88; Conservative 92; Mismatches 176; Indels 119; Gaps 19;  
QY 10 IEYISK-SGKTQENRNGSIGPSIVCKSIQMNQNAENSIQEQEGPLDLTVNRMQEQNTQOG 68

DB 533 LQYLPKDSGDMACRKAASFEKITTSLOOQLOKSEKELVQS-FELKIALEEQ----- 586  
QY 69 DGVLDLSTKTKTSIKSESSICDPSSNSVAGRLHNRNEDYDYS--ASFADGLISKALKD 126  
DB 587 -----LSVKAKNLEMTVNSREHSINAEVQTDVKEVVRKMSVLGDSGYNASND 636  
QY 127 IQSGALDINKAGILYG-----IPQKTLHL--LEALPAGKASFKNKT--RDFHDSY 174  
DB 637 LQSSVDGKRLSSSHDECIHRKMLEQKIVDLEEFLENLK-KSENDKQKSEQQDFMESI 695  
QY 175 SYKDS--KETCAVLQKVALW-----ARQAERTKSKLNLETS 211  
DB 696 QLCEAIIAEMAANALBELALMRDNFDNIILENETLKREIADLERSUKENQETNEFEILE-K 754  
QY 212 EIKFTTASTYLHOL--TLQKMTQFKEKNELOYESNPTVOLKIPOLRVSSVSK----- 264  
DB 755 ETQKEHEAQLIHEITGSLKLVENAEYMNQNLLEDETETKLLKEQEIQLAELRKADNLQ 814  
QY 265 -----SQPDGSGLLDMVYQVSKTSVLEGSALQKLNILPKONKIEGSPVTHSS 314  
DB 815 KKVRFNFDLSVSGMDSEKLECEIFOLKQSLSDAEA-----VTRDAQKECS----- 858  
QY 315 VDSYFLHGDLSPLCLNSKNGTVGTSNTEDGLDRK-----SKOPKRGRIYQYDHE 368  
DB 859 -----FLRSENLELKEKVEDTSMNYNOKERAAASLEFKEOLETEKSNYKKMEAD 905  
QY 369 IMEEAIAVMWSGKMSVSKAGGIYGV--VPHS-----TLEYKVKERSGTLPKPKK 416  
DB 906 LOKE-----LQSAFNEINYINGLAGKVPRLDLSRVELEKKEVSEFKQLEKALEEK 956

RESULT 14  
JCS837  
36K Golgi complex-associated protein - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 05-Mar-1998 #sequence\_revision 13-Mar-1998 #text\_change 09-Jul-2004  
C:Accession: JCS837  
R:Tokii, C.; Fujiwara, T.; Sohma, M.; Hong, H.S.; Misumi, Y.; Ikehara, Y.  
Cell Struct. Funct. 22, 565-577, 1997  
A:Title: Identification and characterization of rat 364-kDa Golgi-associated protein rec  
A:Reference number: JCS837; MUID:98093490; PMID:9431462  
A:Accession: JCS837  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-3187 <TOR>  
A:Cross-references: UNIPROT:O637114; DBJ:D25543; NID:G516825; PIDN:BAA05026.1; PID:G51682  
C:Comment: This protein plays a role in the formation and maintenance of the characterist  
C:Superfamily: giantin  
F:49-549,624-1176,1238-1707,1763-3114/Domain: coiled-coil leucine zipper #status predicte  
F:3165-3187/Domain: membrane anchor #status predicted <NAD>  
Query Match 5.5%; Score 124.5; DB 2; Length 3187;  
Best Local Similarity 21.3%; Pred. No. 23;  
Matches 85; Conservative 64; Mismatches 129; Indels 121; Gaps 20;  
QY 34 KSIQMNQNAENSIQEQEGPLDLTVNRMQEN-----TQQDGVLDLSTKTSIKSESS 86  
DB 2197 KEIWSKAQTELQHQOK-----AYDKLQENKELMSQLEAGQLYHDSKNELTKLESELK 2251  
QY 87 SICDPSS--NSVAG-RLHNRNEDYDYSASFADGLSKALXDIQSGALDINKAGILYGI 143  
DB 2252 SLKQDSTDLKNSLEKCREHNN-----LEGIIKQOEADIQN----- 2287  
QY 144 PQKTLHLLEALPAGKASFKNTRDFHDSYKDSKETCAVLQKVALWARAQAERTKS 203  
DB 2288 -----CKFNCEQLETDLTASRELTTTR-LHDEINVKE-----QKI----- 2320  
QY 204 KNLLETSEIKFTTASTYLHQ-----LTLQKMTQFKEKNELOYESNPTVOLKIPOL 257  
DB 2321 -ISLLSGKEAIIQVIAELHQHQSKEIKLENLLSQ--EEENLTLEENKRAVEKTNOL 2377  
QY 258 R-----VSSVSKSQPDGSGLLDMVYQVSK--TSSVLEGSALQK 295

```

Db      2378 TEALETIKESLSQKALDSFVKSMSLQDRDRIVSDYQLEERHLSVILEKDEL--IQ 2435
QY      296 NILPKQNKI--ECGSPVTHSSVDSYFLHGDLSPLCLNSKNGTVDGT----SENTEDGLDR 349
Db      2436 DAAAEENKLEKEIRG-----LRGHMDD--LNSENAKLDAELIQYRRDLNEVITI 2482
QY      350 KDSKQPRKGRYQYDHEIMEBAIAMVMGSKMSVSKAQ 388
Db      2483 KDSQORQLLEAQLQQ-NKELNECVK--LEGRLKGSSEAE 2518

RESULT 15
Ti6270
hypothetical protein F35D11.11 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: T16270
R:Fulton, B.
submitted to the EMBL Data Library, June 1995
A:Description: The sequence of C. elegans cosmid F35D11.
A:Reference number: Z18487
A:Accession: T16270
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1827 <FUL>
A:Cross-references: UNIPROT:Q20042; EMBL:U29381; NID:g868214; PID:g868224; PIDN:AAA68757
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP.F35D11.11
A:Introns: 76/2; 131/3; 159/3; 185/3; 221/3; 253/3; 320/1; 369/3; 1133/3; 1205/2; 1250/1

Query Match      5.5%; Score 123.5; DB 2; Length 1827;
Best Local Similarity 19.6%; Pred. No. 13;
Matches 92; Conservative 71; Mismatches 206; Indels 101; Gaps 17;

QY      18 KTOENRNGSIGPVSIVCKSIQMQAENSLOB--EQEGPLDLTVNRM-----OEQNTQQG 68
Db      1028 QNSELKNKREG-----LSEKNNEERKKIQDLADQLREANKVHNMRMKNVNLEEKNELD 1082
QY      69 DGVLDSLSTKTSIKSESSICDPSSNSVAGRLHRNREDYVERSABFADGLLSKALKDIQ 128
Db      1083 QNVTDLTNK--VRQLEIQLMDKAAKNVSGDLLRKN-----HDAQSMLKQAQNE-Q 1131
QY      129 SGALDINKAGILYGIPOKTLILLHLEALPAG--KPASFKNKTDFHDSYSYKDSKETCAV 185
Db      1132 FRITDLEKVRKALQDENQRLVNDLATVKAFAEVKRETSKSAISDILLDKYRSAEKANKGE 1191
QY      186 LQKVAL-----WARAQARTS---KSKNLLETSEIKPTASTYLLHQL-----TLQKMV 231
Db      1192 LQNQLRSLDLATVTLKLEQELKAKDSNRLDSOKRFEVEQSKLANLQKSAVESLQNP 1251
QY      232 TQPKENESIQY-----ETSNPTVQLKIPQLRVSSVSXKQPDGSGLLDVMYQV 279
Db      1252 SSNSRQNRSIYVDIPRAASSIGNENSDVPLRSPSVRFADSSQNRQRAVDSMDVSSV 1311
QY      280 SKTSSVLE-----GSALQKLKNIPLKQNKIECGPVTSHSSVDSYFLHGDLSLSP 326
Db      1312 GVTLRFLKERIEQLEADNADLSDALEKAKDELQQRNEKLADQRMVIERVQLVH----- 1366
QY      327 LCINSKNGTVDG--TSEN-----TEDGLDRKDSQPRKGRYQYDHEIMEEAIAMVMSG 380
Db      1367 --ITEERNTIENMTSQROYLTNEESSRSREHRSRKARISTLEHLHREKESKLAHLR 1424
QY      381 KMSVSKAQGIYGVPHSTLEYKVKERSGTLKTPPKKLRPLDPTGLYNMTDS 430
Db      1425 K-----EIEVLHGQLHDALESKEKA-----TGLVGVDQS 1453
```

**THIS PAGE LEFT BLANK**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 24, 2005, 01:33:28 ; Search time 171 Seconds  
(without alignments)  
1323.620 Million cell updates/sec

Title: US-10-016-768A-8

Perfect score: 2250

Sequence: 1 MKMKIRQFAIEYISKSGKTQ.....GLYNMTDGTGCKNSKPKV 442

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 03.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2114.5	94.0	517	2	Q8CJG4 mus musculus
2	1090	48.4	213	2	Q96NKK1 mus sapien
3	1089	48.4	393	2	Q8C9J6 mus musculus
4	502	22.3	271	2	Q8N3X6 homo sapien
5	501.5	22.3	433	2	Q8BGT2 mus musculus
6	501.5	22.3	531	2	Q6ZFI3 mus musculus
7	499.5	22.2	572	2	Q96JN0 mus sapien
8	498	22.1	406	2	Q7Z723 homo sapien
9	497.5	22.1	433	2	Q8N3L6 homo sapien
10	496.5	22.1	433	2	Q86T33 homo sapien
11	470	20.9	396	2	Q8C9Q0 mus musculus
12	456.5	20.3	380	2	Q80VA8 mus musculus
13	444	19.7	223	2	Q8C9B1 mus musculus
14	444	19.7	315	2	Q8BRN7 mus musculus
15	444	19.7	315	2	Q8BRT8 mus musculus
16	353.5	15.7	200	2	Q86T32 homo sapien
17	288	12.8	219	2	Q6NZG1 mus musculus
18	200.5	8.9	1165	2	Q9VDE0 drosophila
19	200.5	8.9	1188	2	Q7YU18 drosophila
20	193.5	8.6	1598	2	Q95YM8 apis mellif
21	183	8.1	185	2	Q22051 caenorhabdi
22	141	6.3	689	2	Q9FNZ7 oryza sativ
23	133.5	5.9	1109	1	Q00756 oryctolagus
24	133.5	5.9	2492	1	Q9W6V0 pcan troglod
25	133.5	5.9	3616	2	Q9W6V0 gallus gall
26	131.5	5.8	997	1	BIR1 SCHPO
27	131.5	5.8	2492	1	ATRX HUMAN
28	131	5.8	1296	1	BXG CLOBO
29	131	5.8	1489	2	Q63ZU6 xenopus lae
30	130.5	5.8	678	1	HOOK DROVI
31	130	5.8	1820	2	Q75XN5 drosophila
32	129.5	5.8	545	2	017584 caenorhabdi
33	129.5	5.8	1591	2	P97868 mus musculu
34	129.5	5.8	2152	2	Q8MJ06 papio hamad
35	129.5	5.8	3099	2	Q86B12 dictyosteli
36	129	5.7	1748	2	Q6CYG5 kluyveromyc
37	128	5.7	675	2	Q6VRS4 helicobacte
38	128	5.7	1256	2	Q22126 caenorhabdi
39	127	5.6	534	2	Q06631 saccharomyc
40	127	5.6	534	2	Q66RD8 dictyosteli
41	127	5.6	1781	2	Q86KX8 dictyosteli
42	126.5	5.6	951	2	Q6CT81 kluyveromyc
43	126.5	5.6	983	2	Q69530 human herpe
44	126.5	5.6	1066	2	Q6BHG0 debaryomyc
45	126.5	5.6	1073	2	Q6FD39 acinetobact

#### ALIGNMENTS

##### RESULT 1

Q8CJG4 ID Q8CJG4 PRELIMINARY; PRT; 517 AA.

AC Q8CJG4;

DT 01-MAR-2003 (Tremblrel. 23, Created)

DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)

DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)

DE Transcription factor MLR1.

GN Name=MLR1; Synonyms=mlr1;

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RX MEDLINE=22448616; PubMed=12560079; DOI=10.1016/S0014-5793(02)03858-9;

RA Kunieda T., Park J.M., Takeuchi H., Kubo T.;

RT "Identification and characterization of Mir1.2: two mouse homologues

RL of Mblk-1, a transcription factor from the honeybee brain.";

RL PDBS Lett. 535:61-65(2003).

DR EMBL; AB076078; BAC20954.1; -

DR MGD; MGI:2651932; Mlr1.

DR GO; GO:0005634; C:nucleus; IC.

DR GO; GO:0003702; P:RNA polymerase II transcription factor acti. . ; IDA.

DR GO; GO:0006366; P:transcription from Pol II promoter; IDA.

DR InterPro; IPR007889; HTH\_psq.

DR InterPro; IPR011526; HTH\_psq\_like.

DR Pfam; PF05225; HTH\_psq; 1.

DR PROSITE; PS00960; HTH\_PSQ; 1.

SQ SEQUENCE 517 AA; 57316 MW; C97403D3D296C52E CRC64;

Query Match 94.0%; Score 2114.5; DB 2; Length 517;

Best Local Similarity 94.1%; Pred. No. 91e-119;

Matches 416; Conservative 12; Mismatches 13; Indels 1; Gaps 1;

QY 1 MKMKIRQFAIEYISKSGKTQENRNGSIGPSIVCKSIQMNQANSLQAEQEGPLDVTNRM 60

Db 77 MKMKIRQFAIEYISKSGKTQENRNGSIGASLVCKSIQMNQANCLQAEQEGPLDVTVRT 136

QY 61 QEQTQGGVLDLSTKTSIKSEESSICDPSSNSVAGLHNRNREYVERSAEFADGLL 120

Db 137 QEQTQGGVLDLSTKTSIKSEESSICDPSSNSVAGLHNRNREYVERSAEFADGLL 196

QY 121 SKALKDIQSGALDINKAGILYGPQKTLHLBALPAGKPAFPKNTDRPHDSYSYKDSK 180

Db 197 SKALKDIQSGALDINKAGILYGPQKTLHLBALPAGKPAFPKNTDRFNDSYSYNEK 256

QY 181 ETCAVLQKVALWARAQARTKSKLNLETSEIKFTASTYLHQLTLQKMWTFKKNES 240

Db 257 ETCAVLQKVALWARAQARTKSKLNLETSEIKFTASTYLHQLTLQKMWTFKKNES 316

QY 241 LQVETSNPTVQLKIPQIRVSSSVKSPQDGSGLLDVMYQVSKTSVLEGSLQKLNLPK 300

Db 241 LQVETSNPTVQLKIPQIRVSSSVKSPQDGSGLLDVMYQVSKTSVLEGSLQKLNLPK 300

Matches 212; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 230 MVTQFKENSLQVETSNPTVQLKIPOLRVNSVSKSQDGSGLLDVWYQVSKTSSVLEGSALQKLNILPK 376

Db 301 QNKIECSGPTVTHSSVDSYFLHGDLSPLCLNSKNGTVDGTSNTEDGLDRKDSKQPRKRG 360

QY 377 QNKLDSCGPTVTHSSVDSYFLHGDLSPLCLNSKNGTVDGTSNTEDGLDRKDKNQKPRKRG 436

Db 361 RYRQYDHEIMEEAIAMVMSGKMSVSKAQGIYGVPHSTLTKYKVKERSGTLKTPPKKRLP 420

QY 437 RYRQYDHEIMEEAIAMVMSGKMSVSKAQGIYGVPHSTLTKYKVKERSGTLKTPPKKRLP 496

QY 421 DTGLYNMTDGTGSCNKSXPV 442

Db 497 DTGLY-MTDSGTGSCNKSXPV 517

RESULT 2

Q96NKL PRELIMINARY; PRT; 213 AA.

ID Q96NKL

AC Q96NKL

DT 01-DEC-2001 (Tremblrel. 19, Created)

DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)

DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)

DE Hypothetical protein FLJ30696.

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RC PubMed=14702039; DOI=10.1038/ng1285;

RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,

RA Wakamatsu A., Hayashi K., Sato H., Tanaka K., Makita H.,

RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,

RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,

RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,

RA Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,

RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,

RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,

RA Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,

RA Tanai H., Kimata M., Watanabe M., Hiraoa K., Chiba Y., Ishida S.,

RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J.,

RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,

RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,

RA Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuma S.,

RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,

RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,

RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,

RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,

RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,

RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,

RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiro M., Ohmori Y.,

RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,

RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigetani S., Senba T.,

RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,

RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,

RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,

RA Okumura K., Negae T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,

RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,

RT "Complete sequencing and characterization of 21,243 full-length human

RT cDNAs";

RL Nat. Genet. 36:40-45(2004).

DR EMBL; AK055258; BAB70892.1; --

DR GO; GO:0005634; C:nucleus; IEA.

DR GO; GO:0003677; F:DNA binding; IEA.

DR InterPro; IPR007889; HTH\_psq.

DR InterPro; IPR011526; HTH\_psq\_like.

DR Pfam; PF05225; HTH\_psq; 1.

DR PROSITE; PS00960; HTH\_PSQ; 1.

SQ SEQUENCE 213 AA; 23477 MW; 4D7F6CABF95251B2 CRC64;

QY 230 MVTQFKENSLQVETSNPTVQLKIPOLRVNSVSKSQDGSGLLDVWYQVSKTSSVLEGSALQKLNILPK 376

Db 301 QNKIECSGPTVTHSSVDSYFLHGDLSPLCLNSKNGTVDGTSNTEDGLDRKDSKQPRKRG 360

QY 377 QNKLDSCGPTVTHSSVDSYFLHGDLSPLCLNSKNGTVDGTSNTEDGLDRKDKNQKPRKRG 436

Db 361 RYRQYDHEIMEEAIAMVMSGKMSVSKAQGIYGVPHSTLTKYKVKERSGTLKTPPKKRLP 420

QY 437 RYRQYDHEIMEEAIAMVMSGKMSVSKAQGIYGVPHSTLTKYKVKERSGTLKTPPKKRLP 496

QY 421 DTGLYNMTDGTGSCNKSXPV 442

Db 497 DTGLY-MTDSGTGSCNKSXPV 517

RESULT 3

Q8C9J6 PRELIMINARY; PRT; 393 AA.

ID Q8C9J6

AC Q8C9J6

DT 01-MAR-2003 (Tremblrel. 23, Created)

DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)

DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)

DE Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched

DE library, clone:A630051A18 product:hypothetical protein, full insert

DE sequence. (Fragment).

GN Name=Mirl;

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Thymus;

RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;

RA Carninci P., Hayashizaki Y.;

RA "High-efficiency full-length cDNA cloning.";

RL Meth. Enzymol. 303:19-44(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Thymus;

RC MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;

RA RIKEN FANTOM Consortium;

RA "Functional annotation of a full-length mouse cDNA collection.";

RL Nature 409:685-690(2001).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Thymus;

RC The FANTOM Consortium,

RA the RIKEN Genome Exploration Research Group Phase I & II Team;

RA "Analysis of the mouse transcriptome based on functional annotation of

RT 60,770 full-length cDNAs.";

RL Nature 420:563-573(2002).

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Thymus;

RC MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;

RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,

RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;

RT "Normalization and subtraction of cap-trapper-selected cDNAs to

RT prepare full-length cDNA libraries for rapid discovery of new genes.";

RL Genome Res. 10:1617-1630(2000).

RN [5]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Thymus;

RC MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;

RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,

RA Konno H., Akiyama J., Nishi K., Kusunagi T., Tashiro H., Itoh M.,

RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,

RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,

Query Match 48.4%; Score 1090; DB 2; Length 213;

Best Local Similarity 99.5%; Pred. No. 7.9e-58;

RESULT 4	PRELIMINARY;	PRT;	271 AA.
Q8N3X6	Q8N3X6		
AC	Q8N3X6;		
DT	01-OCT-2002 (T-EMBLrel. 22, Created)		
DT	01-MAR-2004 (T-EMBLrel. 26, Last sequence update)		
DT	01-MAR-2004 (T-EMBLrel. 26, Last annotation update)		
DE	Hypothetical protein.		
OS	Homo sapiens (human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
QC	NCBI_TaxID=9606;		
RN	[1]		
RC	SEQUENCE FROM N.A.		
RP	TS33UE=Brain;		
RX	MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;		
RA	Straube R.L., Feingold E.A., Grouse L.H., Dege J.G.,		
RA	Klausner R.D., Collins F.S., Wagner L., Shennen C.M., Schuler G.D.,		
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,		
RA	Hopkins R.F., Jordan H., Moore T., Max S.L., Wang J., Hsieh F.,		
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,		
RA	Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,		

## RESULT 5

[4]  
 RP STRAIN=C57BL/6J; TISSUE=Aorta and vein;  
 RA THE FANTOM Consortium;  
 RA THE RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 60,770 full-length cDNAs";  
 RL Nature 420:563-573 (2002).  
 [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;  
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;  
 RA Carninci P., Shibata Y., Hayatsu M., Sugahara Y., Shibata K., Itoh M.,  
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes";  
 RL Genome Res. 10:1617-1630 (2000).  
 [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;  
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Konno H., Akiyama J., Nishi K., Kitsuunai T., Tashiro H., Itoh M.,  
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-Format  
 sequencing pipeline with 384 multicapillary sequencer";  
 RL Genome Res. 10:1757-1771 (2000).  
 [7]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;  
 RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,  
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,  
 RA Hayaehida K., Hayatsu M., Hiramoto K., Hiraoka T., Hirozane T.,  
 RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,  
 RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Koyda M., Koya S.,  
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,  
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,  
 RA Saito R., Saitoh H., Sakai K., Sakai K., Sakazume N., Sano H.,  
 RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,  
 RA Tagawa A., Takahashi F., Takaku-Akashira S., Takeda Y., Tanaka T.,  
 RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB076079; BAC20955.1; -;  
 DR EMBL; AK041090; BAC30816.1; -;  
 DR MGI; MGI:2443930; Mlr2.  
 DR GO; GO:0005634; C:nucleus; IC.  
 DR GO; GO:0003702; F:RNA polymerase II transcription factor acti. . . ; IDA.  
 DR GO; GO:0006366; P:transcription from Pol II promoter; IDA.  
 DR InterPro; IPR007889; HTH\_psq.  
 DR InterPro; IPR011526; HTH\_psq\_like.  
 DR Pfam; PF05225; HTH\_psq; 1.  
 DR PROSITE; PS0960; HTH\_PSQ; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 433 AA; 47124 MW; 736656D1F7E9A041 CRC64;  
 Query Match 22.3%; Score 501.5; DB 2; Length 433;  
 Best Local Similarity 34.5%; Pred. No. 4.3e-22;  
 Matches 162; Conservative 62; Mismatches 143; Indels 103; Gaps 22;  
 QY 1 MKKTIQPAIEYISKSGTKQ-----NRNGS-----IGPSIVCKSIQMNQAEISLQEE 48  
 DB 1 MORMIQOFAEYTSKTSSTQDPNQSTKNQSLPKASPVTTSTAATQNPVLSKLLMAD 60  
 QY 49 QEGPLDLTVNRMQEONTQGGVLDLSTKK-----TSIKSEESSICDPSSSENSVAGRLHRN 104  
 DB 61 QDSPLDLTVRKSSQSEPSQ-DGVLDLSTKKSPCASSTLSHSPGCSSTQGNRGRPSQY 119  
 QY 105 REDIYVERSAEFADGLLSKALKD-----IQSGALDINKAGILYGIPOKTLHLLEALPAG 158  
 DB 218 RPD-----GLRSGDGVPPRSIQDGTREGFGHSTSLKVPLA-----RSLQISELLSRN 265

DB 120 RPD-----GLRSGDGVPPRSIQDGTREGFGHSTSLKVPLA-----RSLQISELLSRN 167  
 QY 159 K-----PASFKNKTDRFDHSYSYKDSKTCVAVLQKVALWARAQAE-RTEKSKLN--- 206  
 DB 168 QLSSTAASLSPGSLQ-----HGQH-----LILSREASWAKPHYEFSLSRMKFRGNG 213  
 QY 207 -LLTSEIKFTFASVYHLQTLQKMTQFKKNSLQYETSNPTVOLKIPOLRVSSYSK 265  
 DB 214 ALSNISDLPLAENS-----APFKVAHQTKQDGRK-DMSHSP-VDLKIPQVRGMOLSW 266  
 QY 266 QPDGSGLLDVMYQVSKTSSVL-----EGSALOKNLNLPKONKIEC--SGPVTHSSVDSYF 319  
 DB 267 SRTGD-----QYSYSLVMGMSQTESALSALLRAILPKQNRKMLDAGP-----DSWG 313  
 QY 320 LHGDLSPCLNSKNGTVDGTSENTEDGLDRKDSQPKRKRGRYQYDHEIMEEIAIAMS 379  
 DB 314 SDAE-----QSTSQPPTSDQEGD-----PGSQPKRKRGRYQYDHEIMEEIAIAMS 363  
 QY 380 GRMSVSKAQGIYGVPHSTLEYKVKERSGTLTKPPKKLRL-----PDTGL 424  
 DB 364 GRMSVSKAQSIYGVPHSTLEYKVKERLGTLPKPPKKMLRSEGPVDSV 413  
 RESULT 6  
 Q6ZPI3 PRELIMINARY; PRT; 531 AA.  
 AC Q6ZPI3;  
 DT 05-JUL-2004 (Tremblrel. 27, Created)  
 DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)  
 DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)  
 DE MKIAA1795 protein (Fragment).  
 GN Name=mkIAA1795;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC TISSUE=Embryonic tail;  
 RX PubMed=14621295;  
 RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hiraoka S.,  
 RA Saga Y., Nagase T., Ohara O., Koga H.;  
 RT Prediction of the coding sequences of mouse homologues of KIAA gene:  
 RT III. the complete nucleotide sequences of 500 mouse KIAA-homologous  
 RT cDNAs identified by screening of terminal sequences of cDNA clones  
 RT randomly sampled from size-fractionated libraries.";  
 RL DNA Res. 10:167-180 (2003).  
 DR EMBL; AK129442; BAC98252.1; -;  
 DR GO; GO:0005634; C:nucleus; IC.  
 DR GO; GO:0003702; F:RNA polymerase II transcription factor acti. . . ; IDA.  
 DR GO; GO:0006366; P:transcription from Pol II promoter; IDA.  
 DR InterPro; IPR007889; HTH\_psq.  
 DR InterPro; IPR011526; HTH\_psq\_like.  
 DR Pfam; PF05225; HTH\_psq; 1.  
 DR PROSITE; PS0960; HTH\_PSQ; 1.  
 FT NON\_TER 1  
 SQ SEQUENCE 531 AA; 58235 MW; EA97247C413A403B CRC64;  
 Query Match 22.3%; Score 501.5; DB 2; Length 531;  
 Best Local Similarity 34.5%; Pred. No. 5.7e-22;  
 Matches 162; Conservative 62; Mismatches 143; Indels 103; Gaps 22;  
 QY 1 MKKTIQPAIEYISKSGTKQ-----NRNGS-----IGPSIVCKSIQMNQAEISLQEE 48  
 DB 99 MORMIQOFAEYTSKTSSTQDPNQSTKNQSLPKASPVTTSTAATQNPVLSKLLMAD 158  
 QY 49 QEGPLDLTVNRMQEONTQGGVLDLSTKK-----TSIKSEESSICDPSSSENSVAGRLHRN 104  
 DB 159 QDSPLDLTVRKSSQSEPSQ-DGVLDLSTKKSPCASSTLSHSPGCSSTQGNRGRPSQY 217  
 QY 105 REDIYVERSAEFADGLLSKALKD-----IQSGALDINKAGILYGIPOKTLHLLEALPAG 158  
 DB 218 RPD-----GLRSGDGVPPRSIQDGTREGFGHSTSLKVPLA-----RSLQISELLSRN 265

Qy	207	-LLETSEIKFPTASTYIHLQTLQKMWTFQFKENSLQVETSNPTVOLKIPOLRVSSVS	265
Db	353	ALGNISDLPLFAENSAFFKMAIQ-----AKQDGKKDVSHSSPYDLKIPQVRGMDLSWE	405
Qy	266	QPGSGLLDVMYQVSKTSSVL-----EGSALQKLNILPKQNKIEC--SGPVTHSSVDVSF	319
Db	406	SRTGDD-----QYSYSSLVWGSQTESALSKKLRAILPRQKSKMLDAGP-----DSWG	452
Qy	320	LHGDSPLCLNSKNGTVDGTSENTEDGLDRKDSQPRKKRGYRYDHEIMEEAIAMVMS	379
Db	453	SDAE-----QSTSGQPYPTSDQSD-----PGSQPRKKRGYRYNSIIEEALSVMS	502
Qy	380	GKMSVSKAQIYGVPHSTLEYKVKVERSGTLTKPKKKLURL-----PDTGL	424
Db	503	GKMSVSKAQSIYGIPIHSTLEYKVKERLGTLLKNPPKKKKMLMRSEGPDSV	552

RESULT 8

Q7Z723

PRELIMINARY; PRT; 406 AA.

AC Q7Z723

DT 01-OCT-2003 (TrEMBLrel. 25, Created)

DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE MUR2 protein.

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OC NCBI\_TaxId=9606;

RP [1]

RN SEQUENCE FROM N.A.

RC TISSUE=Lung;

RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Ustin T.B., Toshikiyuki S., Carninci P., Prange C.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Haile S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Faney J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickens M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,

RA Krzywinski M.I., Skaleka U., Smaluk D.E., Schnerch A., Schein J.E.,

RA Jones S.J., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human

RT and mouse cDNA sequences."

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RP [2]

RN SEQUENCE FROM N.A.

RC TISSUE=Lung;

RA Strausberg R.;

RL Submitted (JUN-2003) to the EMBL/GenBank/DBSJ databases.

DR EMBL; BC053359; AAH53359.1; -.

DR GO; GO:0005634; C:nucleus; IEA.

DR GO; GO:0003677; F:DNA binding; IEA.

DR InterPro; IPR007889; HTH\_psq

DR InterPro; IPR011526; HTH\_psq\_like.

DR Pfam; PF05225; HTH\_psq; 1.

DR PROSITE; PS0960; HTH\_PSQ; 1.

SR SEQUENCE 406 AA; 44083 MW; ACE760CF3CB1F2EF CRC64;

Query Match 22.1%; Score 498; DB 2; Length 406;

Best Local Similarity 33.9%; Pred. No. 6.4e-22;

Matches 156; Conservative 65; Mismatches 141; Indels 98; Gaps 19

Qy	1	MKKWIRFOALEYTKSGKTQE-----NRNGS-----IGPSIVCKSIQMNAENSLQOE	48
----	---	--	----

```
Db 1 MORMIPOFAEYTSKNSSTQDPSPNSTKNQSLPKASPVTTSTPTAATQNPVLSKLLMAD 60
QY 49 QEGPLDITVNRMOEQNTQCGDGLDLSKTCT-----SIKSESSICDPSSNSVAGRLHRN 104
Db 61 QDSPLDITVRKQSEPSSEQ-DGVLDTLSTKSPCAGSTLSHSPGSCSTQNGRPGPSQY 119
QY 105 REDYVERSAEFAADGLLSKALKD-----IQSGALDINKAGILYIGIPQKTLILLHLEALPAG 158
Db 120 RPD-----GLRSGDGVPPRSIQDGTREGFGHSTSLKVPLA-----RSIQISELLSRN 167
QY 159 K-----PASFKNKTRDFHDSYSYKDSKETCAVLQKVALWARAQAE-RTEKSKLN--- 206
Db 168 QLSTAASLGPGLQN-----HGQH-----LILSREASWAKPHYFNLSRMKFRNG 213
QY 207 -LLETSEIKFTASTYHLHTLQKMTQPKKESLQYETSNPTVOLKIPQLRVS SVSKS 265
Db 214 ALSNIDLPFLAENSAFFKMAQ-----AKQDGKDVSHSPVDLKIPIQVRGMDLSWE 266
QY 159 K-----PASFKNKTRDFHDSYSYKDSKETCAVLQKVALWARAQAE-RTEKSKLN--- 206
Db 168 QLSTAASLGPGLQN-----HGQH-----LILSREASWAKPHYFNLSRMKFRNG 213
QY 207 -LLETSEIKFTASTYHLHTLQKMTQPKKESLQYETSNPTVOLKIPQLRVS SVSKS 265
Db 214 ALSNIDLPFLAENSAFFKMAQ-----AKQDGKDVSHSPVDLKIPIQVRGMDLSWE 266
QY 266 QPDGSGLLDWMYQVSKTSSVL-----EGSALQKLKNILPKONKIEC--SGPVTHSSVDSYF 319
Db 267 SRTGD-----QYSYSSLVMGQTESALSCLKRALILPKSRKSMLDAGP-----DSWG 313
QY 320 LHGDLSPLCLNSKNGTVDTGTSNTEDGLDRKQKPKRGYRYQYDHEIMEEAIAMWS 379
Db 314 SDAE-----QSTSGOPYPTSDQEGD-----PGSKQPKRGYRYQYNSILEEAIAMWS 363
QY 380 GKMSVSKAQGIYGVPHSTLEYKVKERSGTLKTPPKKKLRL 419
Db 364 GKMSVSKAQSIYGIPIHSTLEYKVKERLGTILKNPPKKMKML 403

RESULT 9
Q8N3L6 PRELIMINARY; PRT; 433 AA.
ID Q8N3L6 Q8N3L6
AC Q8N3L6 Q8N3L6
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Hypothetical protein DKFZp451A142.
GN Name=DKFZp451A142;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Human skeletal muscle;
RA Wambutt R., Heubner D., Mewes H.W., Weil B., Amid C., Osanger A.,
RA Fobo G., Han M., Wiemann S.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL834245; CAD38921.2; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR InterPro; IPR007889; HTH_psq.
DR InterPro; IPR011526; HTH_psq_like.
DR Pfam; PF05225; HTH_psq; 1.
DR PROSITE; PS50960; HTH_PSQ; 1.
KW Hypothetical protein.
SQ SEQUENCE 433 AA; 47016 MW; 9EDA693DEAF7A564 CRC64;

Query Match 22.1%; Score 497.5; DB 2; Length 433;
Best Local Similarity 33.4%; Pred. No. 7.4e-22;
Matches 157; Conservative 64; Mismatches 146; Indels 103; Gaps 19;

QY 1 MORMIPOFAEYTSKNSSTQDPSPNSTKNQSLPKASPVTTSTPTAATQNPVLSKLLMAD 60
Db 1 MORMIPOFAEYTSKNSSTQDPSPNSTKNQSLPKASPVTTSTPTAATQNPVLSKLLMAD 60
QY 49 QEGPLDITVNRMOEQNTQCGDGLDLSKTCT-----SIKSESSICDPSSNSVAGRLHRN 104
Db 1 MORMIPOFAEYTSKNSSTQDPSPNSTKNQSLPKASPVTTSTPTAATQNPVLSKLLMAD 60
QY 105 REDYVERSAEFAADGLLSKALKD-----IQSGALDINKAGILYIGIPQKTLILLHLEALPAG 158
Db 120 RPD-----GLRSGDGVPPRSIQDGTREGFGHSTSLKVPLA-----RSIQISELLSRN 167
QY 159 K-----PASFKNKTRDFHDSYSYKDSKETCAVLQKVALWARAQAE-RTEKSKLN--- 206
Db 168 QLSTAASLGPGLQN-----HGQH-----LILSREASWAKPHYFNLSRMKFRNG 213
QY 207 -LLETSEIKFTASTYHLHTLQKMTQPKKESLQYETSNPTVOLKIPQLRVS SVSKS 265
Db 214 ALSNIDLPFLAENSAFFKMAQ-----AKQDGKDVSHSPVDLKIPIQVRGMDLSWE 266
QY 266 QPDGSGLLDWMYQVSKTSSVL-----EGSALQKLKNILPKONKIEC--SGPVTHSSVDSYF 319
Db 267 SRTGD-----QYSYSSLVMGQTESALSCLKRALILPKSRKSMLDAGP-----DSWG 313
QY 320 LHGDLSPLCLNSKNGTVDTGTSNTEDGLDRKQKPKRGYRYQYDHEIMEEAIAMWS 379
Db 314 SDAE-----QSTSGOPYPTSDQEGD-----PGSKQPKRGYRYQYNSILEEAIAMWS 363
QY 380 GKMSVSKAQGIYGVPHSTLEYKVKERSGTLKTPPKKKLRL 419
Db 364 GKMSVSKAQSIYGIPIHSTLEYKVKERLGTILKNPPKKMKML 403
```

```
QY 105 REDYVERSAEFAADGLLSKALKD-----IQSGALDINKAGILYIGIPQKTLILLHLEALPAG 158
Db 120 RPD-----GLRSGDGVPPRSIQDGTREGFGHSTSLKVPLA-----RSIQISELLSRN 167
QY 159 K-----PASFKNKTRDFHDSYSYKDSKETCAVLQKVALWARAQAE-RTEKSKLN--- 206
Db 168 QLSTAASLGPGLQN-----HGQH-----LILSREASWAKPHYFNLSRMKFRNG 213
QY 207 -LLETSEIKFTASTYHLHTLQKMTQPKKESLQYETSNPTVOLKIPQLRVS SVSKS 265
Db 214 ALSNIDLPFLAENSAFFKMAQ-----AKQDGKDVSHSPVDLKIPIQVRGMDLSWE 266
QY 266 QPDGSGLLDWMYQVSKTSSVL-----EGSALQKLKNILPKONKIEC--SGPVTHSSVDSYF 319
Db 267 SRTGD-----QYSYSSLVMGQTESALSCLKRALILPKSRKSMLDAGP-----DSWG 313
QY 320 LHGDLSPLCLNSKNGTVDTGTSNTEDGLDRKQKPKRGYRYQYDHEIMEEAIAMWS 379
Db 314 SDAE-----TPGOPYPTSDQEGDPSKQPKRGYRYQYNSILEEAIAMWS 363
QY 380 GKMSVSKAQGIYGVPHSTLEYKVKERSGTLKTPPKKKLRL-----PDTGL 424
Db 364 GKMSVSKAQSIYGIPIHSTLEYKVKERLGTILKNPPKKMKMLRSEGDVSV 413

RESULT 10
Q86T33 PRELIMINARY; PRT; 433 AA.
ID Q86T33 Q86T33
AC Q86T33 Q86T33
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Hypothetical protein DKFZp313C088.
GN Name=DKFZp313C088;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RG The German cDNA Consortium;
RA Wambutt R., Heubner D., Mewes H.W., Weil B., Amid C., Osanger A.,
RA Fobo G., Han M., Wiemann S.;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL832106; CAD91159.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR InterPro; IPR007889; HTH_psq.
DR InterPro; IPR011526; HTH_psq_like.
DR Pfam; PF05225; HTH_psq; 1.
DR PROSITE; PS50960; HTH_PSQ; 1.
KW Hypothetical protein.
SQ SEQUENCE 433 AA; 46975 MW; 9089929710DF291E CRC64;

Query Match 22.1%; Score 496.5; DB 2; Length 433;
Best Local Similarity 33.6%; Pred. No. 8.5e-22;
Matches 158; Conservative 65; Mismatches 144; Indels 103; Gaps 20;

QY 1 MORMIPOFAEYTSKNSSTQDPSPNSTKNQSLPKASPVTTSTPTAATQNPVLSKLLMAD 60
Db 1 MORMIPOFAEYTSKNSSTQDPSPNSTKNQSLPKASPVTTSTPTAATQNPVLSKLLMAD 60
QY 49 QEGPLDITVNRMOEQNTQCGDGLDLSKTCT-----SIKSESSICDPSSNSVAGRLHRN 104
Db 61 QDSPLDITVRKQSEPSSEQ-DGVLDTLSTKSPCAGSTLSHSPGSCSTQNGRPGPSQY 119
QY 105 REDYVERSAEFAADGLLSKALKD-----IQSGALDINKAGILYIGIPQKTLILLHLEALPAG 158
Db 120 RPD-----GLRSGDGVPPRSIQDGTREGFGHSTSLKVPLA-----RSIQISELLSRN 167
QY 159 K-----PASFKNKTRDFHDSYSYKDSKETCAVLQKVALWARAQAE-RTEKSKLN--- 206
```

```

Db 168 QLSTAAISLSPGLN-----HGQH-----LILSRASWAKPHYFNLSRMKFRNG 213
Qy 207 -LLETSEIKPPTASTYLHQLTQMWVTFQKEKNESLQYTSNPTVLKIPQLRVSSVSKS 265
Db 214 ALSNISDLPLAENSAPPKMALQ-----AKDGKDVSHSHSPVDLKIPOVRGMDLSWE 266
Qy 266 QPDGSGLLDMVYQVSKTSSVL-----EGSALQKLKNILPKQNKIEC--SGPVTHSSVDYSF 319
Db 267 SRTGD-----QYSYSSLVMGQTESALSKKLRAILPKQSRKSMMLDAGP-----DSWG 313
Qy 320 LHGDLSPLCLNSKNGTVDTGSENTEGLDRKDSQPKRGRVROYDHEIMEBAIAVMMS 379
Db 314 SDAE-----QSTSGQPYPTSDQEGD-----PGSKQPKRGRVROYNSEILEBAISVMS 363
Qy 380 GKMSVSKAQSIYGVPHSTLLEYKVKERSGTLKTPPKKRL-----PDTGL 424
Db 364 GKMSVSKAQSIYGVPHSTLLEYKVKERLGLTKNPPKKQKMLRSEGGPDVSV 413

RESULT 11
ID Q8C9Q0 PRELIMINARY; PRT; 396 AA.
AC Q8C9Q0;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DE Mus musculus 3 days neonate thymus cDNA, RIKEN full-length enriched
DE library, clone:A630025C20 product:hypothetical protein, full insert
DE sequence. (Fragment).
GN Name=Mlr2;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RA The FANTOM Consortium.
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,

```

```

RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sobabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK041621; BAC31007.1; --
DR MGD; MGI:2443930; Mlr2.
DR GO; GO:0005634; C:nucleus; IC.
DR GO; GO:0003702; F:RNA polymerase II transcription factor acti. . . ; IDA.
DR GO; GO:0006366; P:transcription from Pol II promoter; IDA.
DR InterPro; IPR007889; HTH_psq.
DR InterPro; IPR011526; HTH_psq_like.
DR Pfam; PF05225; HTH_psq; 1.
DR PROSITE; PS0960; HTH_PSQ; 1.
KW Hypothetical protein.
FT NON_TER 396
SQ SEQUENCE 396 AA; 43085 MW; EE4A585F62336E35 CRC64;

Query Match 20.9%; Score 470; DB 2; Length 396;
Best Local Similarity 34.2%; Pred. No. 2.9e-20;
Matches 155; Conservative 59; Mismatches 141; Indels 98; Gaps 21;

Qy 1 MKKMIQFAIEYISKSGKQEQ-----NRNGS-----IGPSIVCKSIQNNQAEISQEE 48
Db 1 MQEMIQFAAEYTSKTSSTQDPSPNSTKNQSLPKASPVTTSTAATQNPVLSKLLMAD 60

Qy 49 QEGPLDITVNRMQEONTQGGDGVLDLSTKK-----TSIKSEESSICDPSSNSVAGRLHRN 104
Db 61 QDSPLDLTVRKSSQSEPEQ-DGVLDLSTKSPCASSTLSHSPGCSSTQGNRGRPSQY 119

Qy 105 REDYVERSAEFADGLLSKALD-----TQSGALDINKAGILYIPQKTLHLLEALPAG 158
Db 120 RPD-----GLRSGDGVPPRLQDGTREGFGHSTLSKVPLA-----RSIQISELLSRN 167

Qy 159 K-----PASPFKNKTRDPHDSYSYKDSKETCAVLQKVALWARAQAE-RTEKSKLN--- 206
Db 168 QLSTAAISLSPGLN-----HGQH-----LILSRASWAKPHYFNLSRMKFRNG 213

Qy 207 -LLETSEIKPPTASTYLHQLTQMWVTFQKEKNESLQYTSNPTVLKIPQLRVSSVSKS 265
Db 214 ALSNISDLPLAENS-----APPKMAHQTKQDCKR-DMSHSSP-VDLKIPQVRGMDLSWE 266

Qy 266 QPDGSGLLDMVYQVSKTSSVL-----EGSALQKLKNILPKQNKIEC--SGPVTHSSVDYSF 319
Db 267 SRTGD-----QYSYSSLVMGQTESALSKKLRAILPKQSRKSMMLDAGP-----DSWG 313

Qy 320 LHGDLSPLCLNSKNGTVDTGSENTEGLDRKDSQPKRGRVROYDHEIMEBAIAVMMS 379
Db 314 SDAE-----QSTSGQPYPTSDQEGD-----PGSKQPKRGRVROYNSEILEBAISVMS 363

Qy 380 GKMSVSKAQSIYGVPHSTLLEYKVKERSGTLKTP 412
Db 364 GKMSVSKAQSIYGVPHSTLLEYKVKERLGLTKNKP 396

RESULT 12
Q80VA8

```

```

ID Q80VAB PRELIMINARY; PRT; 380 AA.
AC Q80VAB;
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DE 01-MAR-2003 (T-EMBLrel. 26, Last annotation update)
DE M12 protein (fragment).
GN Name=M12;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Mammary tumor; DOI=10.1073/pnas.242603899;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner K.H., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skaleka U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Mammary tumor;
RA Strausberg R.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC050068; AAH50068.1; -.
DR MGD; MGI:2443930; M12.
DR GO; GO:0005634; C:nucleus; IC.
DR GO; GO:0003702; F:RNA polymerase II transcription factor acti. . ; IDA.
DR GO; GO:0006366; P:transcription from Pol II promoter; IDA.
DR InterPro; IPR007889; HTH_psq.
DR InterPro; IPR011526; HTH_psq_like.
DR Pfam; PF05225; HTH_psq; 1.
DR PROSITE; PS50960; HTH_PSQ; 1.
FT NON TER 1
SQ SEQUENCE 380 AA; 41445 MW; 4D3B6820C3B15BEE CRC64;

Query Match 20.3%; Score 456.5; DB 2; Length 380;
Best Local Similarity 35.3%; Pred. No. 1.8e-19;
Matches 146; Conservative 54; Mismatches 123; Indels 91; Gaps 20;

QY 45 LQEQGSLDLTVNRMQEONTQGGDGLDLSTVKK---TSIKSEESICDPSSSENSVAGR 100
DB 4 LMDQSPDLTVRKQSQSEFSEQ-DGVLDTLSTKSCASTSLSHSPGCSSTGNGRPGR 62
QY 101 LHRNREDYVERSAEFADGLSLKALKD-----IQSGALDINKAGILYGPQKTLHLLEA 154
DB 63 PSQVRPD---GLRSGDGVPPRSIQDGTREGFGHSTSLKVPLA-----RSLQISEEL 110
QY 155 LPAGK-----PASFKNKTRFDHDSYSYKDSKETCAVLQKVALWARAQAE-RTEKSKL 205
DB 111 LSRNQLSTRAASLGFSGQLN-----HGQH-----LILSRSAWAKPHYFSLSRMKF 156
QY 206 N-----LLETSEIKFPTATYHLQTLQKMWYQPKKNESLQYETSPTVOLKIPQLRVSS 261
DB 157 RGNGLSNTSDLPFLAENS-----APFKMAHQTKQDKR-DMHSHP-VDLKIPQVRGMD 209
QY 262 VSKSQPDGSGLLDMYQVSKTSSVL-----EGSALQKLNLKPKNKIEC--SGPVTTHSV 315
```

```

DB 210 LSWESRTGD-----QYSYSLVMGSGQTESALSCLKRAILPKQNRKSMLDAGP----- 256
QY 316 DSYFLHGLDLSPLCLNSKNGTVDGTSENTEDGLDKDSQKPRKRGYRYDYDHEIMEEATA 375
DB 257 DSWGSDAE-----QSTGQPYPTSDQEGD-----PGSQPRKKRGYRYNISEILEALS 306
QY 376 MYMSGMSVSKAQGIYGVPHSTLEYKVKERSGTLTTPPKKKLRL-----PDTGL 424
DB 307 VMSGMSVSKAQSIYGIPHSHTLEYKVKERLGTLPKPPKKMKLMRSEGPDVSV 360
NCBI_TaxID=10090;
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/350555500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RA The FANTOM Consortium;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX MEDLINE=2030913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Iehi Y., Nakamura S., Hazama M., Nishine T., Kashiwagi K.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaki S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
```



```

RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori P., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Kato H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Sano H.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.,
RA Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RL ENBL; AK042567; BAC31295.1; -.
KW Hypothetical protein.
SQ SEQUENCE 223 AA; 24472 MW; B019FF8BFCB7C72F CRC64;

Query Match 19.7%; Score 444; DB 2; Length 223;
Best Local Similarity 80.4%; Pred. No. 4.8e-19;
Matches 90; Conservative 7; Mismatches 15; Indels 0; Gaps 0;

QY 1 MKKMIRQFAIEYISKGTQENRNGSIGSVCKSIQMNQAEISLOEQEGPLDLTVNRM 60
Db 77 MKKMIRQFAIEYISKGTQENRNGSIGSVCKSIQMNQAEISLOEQEGPLDLTVNRT 136

QY 61 QEQNTQGGVLDLSTKTKTSIKSESSICDPSSSENSVAGRLHNRREDYVERS 112
Db 137 QEQTAQGGVLDLSTKTKTSIKSESSISDPSSSENSVAGMLQMKTKDEKVDLS 188

RESULT 14
Q8BRT7 PRELIMINARY; PRT; 315 AA.
AC Q8BRT7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Mus musculus 10 days neonate cortex cDNA, RIKEN full-length enriched
DE library, clone:A830039H10 product:hypothetical protein, full insert
DE sequence.
DE Name=Mlrl;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cortex;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cortex;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cortex;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;

```

```

RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cortex;
RX MEDLINE=20530933; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Fujiwako R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Yamanote S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cortex;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori P., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Kato H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR MGD; MGI:2651932; Mlrl.
DR GO; GO:0005634; C:nucleus; IC.
DR GO; GO:0003702; F:RNA polymerase II transcription factor acti. . . ; IDA.
DR GO; GO:0006366; P:transcription from Pol II promoter; IDA.
KW Hypothetical protein.
SQ SEQUENCE 315 AA; 34830 MW; CD99B448E52A650C CRC64;

Query Match 19.7%; Score 444; DB 2; Length 315;
Best Local Similarity 80.4%; Pred. No. 7.8e-19;
Matches 90; Conservative 7; Mismatches 15; Indels 0; Gaps 0;

QY 1 MKKMIRQFAIEYISKGTQENRNGSIGSVCKSIQMNQAEISLOEQEGPLDLTVNRM 60
Db 160 MKKMIRQFAIEYISKGTQENRNGSIGSVCKSIQMNQAEISLOEQEGPLDLTVNRT 219

QY 61 QEQNTQGGVLDLSTKTKTSIKSESSICDPSSSENSVAGRLHNRREDYVERS 112
Db 220 QEQTAQGGVLDLSTKTKTSIKSESSISDPSSSENSVAGMLQMKTKDEKVDLS 271

RESULT 15
Q8BRT8 PRELIMINARY; PRT; 315 AA.
AC Q8BRT8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Mus musculus 10 days neonate cortex cDNA, RIKEN full-length enriched
DE library, clone:A830001M15 product:hypothetical protein, full insert
DE sequence.
DE Name=Mlrl;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cortex;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;

```

